Protein–Protein Interactions in Epidermal Growth Factor Receptors Through Molecular Dynamics

⊢ Juan Felipe Franco-Gonzalez ⊣

Supervised by:
Dr. Victor L. Cruz

2014

Protein–Protein Interactions in Epidermal Growth Factor Receptors Through Molecular Dynamics

⊢ Juan Felipe Franco-Gonzalez ⊣

Supervised by:
Dr. Victor L. Cruz

This thesis has been submitted to obtain the title of Doctor at:

2014

To the reader in search of knowledge or data.
Your readership means this thesis
has been worth the effort.

Resumen

Los Receptores de Factor de Crecimiento Epidérmico (EGFR, en inglés) son esencialmente cuatro proteínas: EGFR/ErbB1, ErbB2, ErbB3 y ErbB4. Éstas están asociadas con determinados procesos biológicos y cada vez son reconocidas como dianas terapéuticas importantes contra el cáncer. En esta tesis, se proporcionan modelos basados en homología para los dominios extracelulares (ectodomio, ECD en inglés) de ErbB3 y ErbB4 en sus conformaciones activas, incluyendo el ligando heregulin, seguido de un posterior refinamiento de los modelos a través de dinámica molecular a una resolución atomística (AA). Un modelo construído para la ErbB2 basado en información cristalográfica permitió el análisis de las características comunes observadas entre los miembros de esta familia, concretamente, el *movimiento de periscopio* del brazo de dimerización y el *movimiento de bisagra* del subdominio IV. Adicionalmente a esta parte, se proporciona un modelo refinado para la interacción en las regiones ECD correspondientes al heterodímero ErbB2/ErbB3. Este heterodímero es ampliamente reconocido por tener un alto impacto en el desarrollo del cáncer.

El receptor ErbB2 es considerado como una onco-proteína transmembranal que está sobre-expresada en el cáncer de mama. Un tratamiento terapéu-tico exitoso está basado en el uso de un anticuerpo monoclonal llamado Trastuzumab el cual se sabe que interactúa con el dominio extracelular (ECD-ErbB2). Un mejor entendimiento de la estructura de forma detallada, en la interacción receptor–anticuerpo es de primordial interés en el diseño de terapias anticáncer más efectivas. Con el fin de discutir la flexibilidad del complejo ECD-ErbB2/Trastuzumab, se han hecho tanto una simulación de dinámica molecular de multi-segundos como una análsis de componentes principales (PCA, en inglés) de este sistema. Con el propósito de validar estas simulaciones, se ha realizado un análisis detallado de las interacciones entre el dominio variable del anticuerpo y el subdominio IV de ECD-ErbB2. Esta estructura ha sido elucidada estáticamente mediante cristalografía de Rayos-X. Ciertamente, los resultados de la simulación están en excelente concordancia con la información experimental disponible durante toda la trayectoria. Los resultados de PCA muestran fluctuaciones colectivas asocia-das a un

movimiento de *bisagra*, en el cual, el subdominio II y el dominio constante (C_H) se acercan entre sí. Este movimiento es probablemente estabilizado por la fomación de puentes de hidrógeno y puentes salinos entre residuos del brazo de dimerización en el subdominio II y residuos localizados en el dominio constante C_H del anticuerpo Trastuzumab. Finalmente, en este punto, se discutió la flexibilidad del modelo descrito por dinámica molecular en relación con la estructura cristalográfica. Se ha reportado por primera vez un movimiento del anticuerpo hacia el dominio de dimerización del receptor ErbB2. Este hallazgo podría tener consecuencias importantes en la acción biológica del anticuerpo monoclonal.

Además de lo anteriormente mencionado en modelos puramente atómicos, se ha validado un mapeo desde la resolución AA hacia una resolución de grano–grueso (*Coarse–Grained*, CG, en inglés). De esta manera, AA MD sobre el complejo ECD-ErbB2/Trastuzumab-Fab ha sido usada para comparar con simulaciones CG MD. Específicamente, se ha comparado el campo de fuerzas CG Martini con el AA OPLS. Se ha analizado la flexibilidad conformacional y las interacciones entre el anticuerpo y el receptor. En esta tesis han sido examinados los siguientes parámetros en los algoritmos de MD para llevar a cabo las simulaciones con Martini: el método de cálculo de interacciones no-enlazantes para la descripción de las interacciones electrostáticas, el valor del radio *cut-off* de la lista de vecinos (r_{list}) y el método de Redes Elásticas (*Elastic Network*, EN, en inglés). Los resultados muestran que en simulaciones de CG MD los modelos basados en *domElNeDyn* (redes elásticas por dominios), PME y un r_{list} de 1.4 nm, son comparables a los modelos AA. Los resultados proporcionan una luz para validar el campo de fuerzas Martini tanto en las interacciones proteína–proteína como en la predicción de estructura.

Siguiendo este formalismo se han aplicado simulaciones CG MD han sido aplicadas para estudiar la influencia del Trastuzumab en la estructura y dinámica del dímero ErbB2 entero, incluyendo la bicapa lipídica. El uso de modelos CG para estudiar tales complejos es al menos de momento, obligatorio, debido al gran tamaño del sistema entero. El modelo Martini lo hace satisfactoriamente bien, arrojando resultados a la par con aquellos obtenidos por modelos AA, como también con la información experimental existente sobre receptores homólogos. Por ejemplo, los dominios ECD e intracelular se aproximan a la superficie de la bicapa lipídica en sendos casos, el monómero y el dímero. El Trastuzumab-Fab dificulta la interacción de los receptores con la bicapa lipídica. Otro efecto interesante del anticuerpo es la alteración de la disposición anti-paralela de los segmentos yuxtamembrana en el caso del dímero. Estos hallazgos ayudan a entender el efecto del anticuerpo sobre la bioactividad del receptor.

Summary

The family of epidermal growth factor receptors (EGFR) is composed by four members: EGFR/ErbB1, ErbB2, ErbB3 and ErbB4. They are associated with a number of biological processes and are becoming increasingly recognized as important therapeutic targets against cancer. In this thesis, some models, based on homology, there are provided for the extracellular domains (ectodomain, ECD) of ErbB3 and ErbB4 in their active conformations, including a heregulin ligand, followed by further refinement of the models by molecular dynamics (MD) simulations at atomistic (AA) resolution. A model built for ErbB2 based on crystallographic information allowed an analysis of the common features observed among members of the family, namely, the periscope movement of the dimerization arm and the hinge displacement of subdomain IV. In addition, a refined model for the interaction of the ECD corresponding to a ErbB2/ErbB3 heterodimer is given. This heterodimer is widely recognized to have a high impact in cancer development.

The ErbB2 receptor is a transmembrane oncoprotein that is over expressed in breast cancer. A successful therapeutic treatment is a monoclonal antibody called Trastuzumab which interacts with the ErbB2 extracellular domain (ECD-ErbB2). A better understanding of the detailed structure of the receptor-antibody interaction is indeed of prime interest for the design of more effective anticancer therapies. To analyze the flexibility of the complex ECD-ErbB2/Trastuzumab, a multi-nanosecond MD simulation together with an analysis of fluctuations through a principal component analysis (PCA) of this system, was carried out. For validating the simulations, a detailed analysis of the variable antibody domain interactions with the extracellular subdomain IV of ErbB2 was performed. This structure has been statically elucidated by X-ray crystallography. Indeed, the simulation results are in excellent agreement with the available experimental information. The PCA shows collective fluctuations resulting in a hinge motion in which subdomain II and constant domain (C_H) approach each other. This movement is likely stabilized by the formation of H-bonds and salt bridge interactions between residues of the dimerization arm in the subdomain II and the

antibody Trastuzumab residues located in the C_H domain. Finally, the flexibility of the molecular dynamics model in relation with the static X-ray structure was discussed. A movement of the antibody towards the dimerization domain of the ErbB2 receptor is reported for the first time. This finding could have relevant consequences on the biological action of the monoclonal antibody.

In addition to the above mentioned pure atomistic models, a mapping from AA to coarse-grained (CG) resolution has been validated. In this manner AA MD on ECD-ErbB2/Trastuzumab-Fab has been used to compare with CG MD simulations. Specifically, the CG Martini force field has been compared with the AA OPLS representation. The conformational flexibility and interactions between the antibody and the receptor have been analyzed. In this thesis the following parameters have been tested in the MD algorithms to carry out the Martini simulations: the non-bonded interactions methods to calculate the electrostatic interactions, the value of the neighbor lists cut-off radius (r_{list}) and the Elastic Network method. The results show that when used in MD simulations domElNeDyn models, PME and an r_{list} of 1.4 nm, are comparable to the AA protein models. The results shed light to validate the Martini force field in the protein–protein interactions and towards protein prediction structure.

Therefore, CG MD simulations have been applied to study the influence of the Trastuzumab monoclonal antibody on the structure and dynamics of the full-length ErbB2 receptor dimer, including the lipid bilayer. The usage of CG models to study such complexes is almost mandatory, at present, due to the large size of the whole system. The Martini model performs satisfactorily well, giving results well-matched with those obtained by AA models as well as with the experimental information existing on homologous receptors. For example, the ecto and intracellular domains approach the bilayer surface in both the monomer and dimer cases. The Trastuzumab-Fab hinders the interaction of the receptors with the lipid bilayer. Another interesting effect of the antibody is the disruption of the antiparallel arrangement of the juxtamembrane segments in the dimer case. These findings might help to understand the effect of the antibody on the receptor bioactivity.

Acknowledgments

This thesis was performed with the financial support of MINECO through the FPI program with the project MAT-200912364 and the scholarship number BES-2010-041260. The computational resources access was given thanks to SGAI-CSIC and CESGA

I would like to thank ...

My thesis supervisor Victor L. Cruz, for his perceptive direction. Victor, thanks to you I joined the Biophym group four year ago. Since the first and unique call made from across the sea you trusted me.

Group's head Pr. Dr. Javier Martinez, for let me this great opportunity to work at Biophym group.

Dr. Javier Ramos, for your helpful discussions.

Dr. Juan F. Vega, you have always been available to help.

Dr. Manuel Alcamí, for accept to be the tutor of this thesis.

Dr. Guillermo Mena, for allow me to work at IEM.

My former professors in Colombia, Cacier Hadad and Fernando Echeverri, for help me to find a PhD scholarship in Spain.

The Martini group's head, Dr. S. J. Marrink, for let me the opportunity to work in his group.

The Martini team, for your valuable and profitable help: Djurre, Xavier and Tsjerk.

Nik, thank you for your helpful and refined corrections.

Rafa, thank you for your useful suggestions. Also, thank you for let me know the hidden places of Madrid to enjoy caña and tapa for a cheap price attached to a detalled historical story and show me the amazing places in *la Sierra* for trekking.

LuzK, thank you for motivate me to *cross the pond*. Thank you for fully share with me the first years of my stay in Madrid.

Noe, thank you for your constant help, support, advices, suggestions and convince me to enjoy every weekend's plan whatever it is. Thank you for convey to me those almost invisible things belong to the routine.

My parents, Juan y Nena, for always be there when I call.
Thanks to my friends and colleages, Santa, Jymmy, Almudena, Jhon Fredy and Mauro.
The funny and dynamic people of Groningen, Clement, Marcelo, Katia, Jaakko, Susan and Thomas.
I can't to forget Teresa Burriel, who was waiting on the line of the Spanish Immigration Service a day at 8 a.m in my first and cold week in Madrid.
Thanks to those beautiful people of both groups *"danzas"* and *"bicis"*, Alvaro, Raúl, Laura, Andoni, Alicia, Laurita, Giada, Chiara, Inma, Clara, Luisón, Pep, María, Luisín, Rubén, Diana, Mariaje, JuanFe and the inequable *"el mata"*.
Also, thanks to the people of *"ecoarganzuela"*, Pacho, Nadia, Maie, Alejo, Mireia and Joaquin.
Thanks to supergirls team *"las chatis"*, Lucía, María and Carla.

Contents

Part I

Introduction and Overview

Chapter 1

Molecular Dynamics Simulations

Molecular Dynamics (MD) can be defined as the study of the molecular movement caused by atomic interactions. However, the term molecular dynamics has become a synonym for molecular dynamics computer simulations. This Molecular Dynamics Subject has been studied since the 19th century. The field of computational molecular dynamics, due to its dependence on extensive, iterated computation, had to wait for the invention of the electronic computer in the mid 20th century to be feasible. Despite this relatively short history, the field has expanded at lightning pace and found applications in a wide range of research areas. Both Levitt [1] and Karplus [2] may be considered the founders of the MD developments on the light of their personal research. Since then, a lot of work has been done to improve the methodology. In an effort to compile these works, some papers provide a combination of a short historical overview and an outlook to the future [3, 4]. This chapter points to some studies that have played an important role in the field treated in this thesis: molecular dynamics simulations of Epidermal Growth Factor Receptors.

1.1 A Brief History

There are a lot of experimental techniques that can provide information about the dynamics of biomolecules, but they are limited in their spatial and temporal resolution. Most of these techniques report ensemble average properties rather than the motion of individual molecules. Some examples can be seen in Figure 1.1.

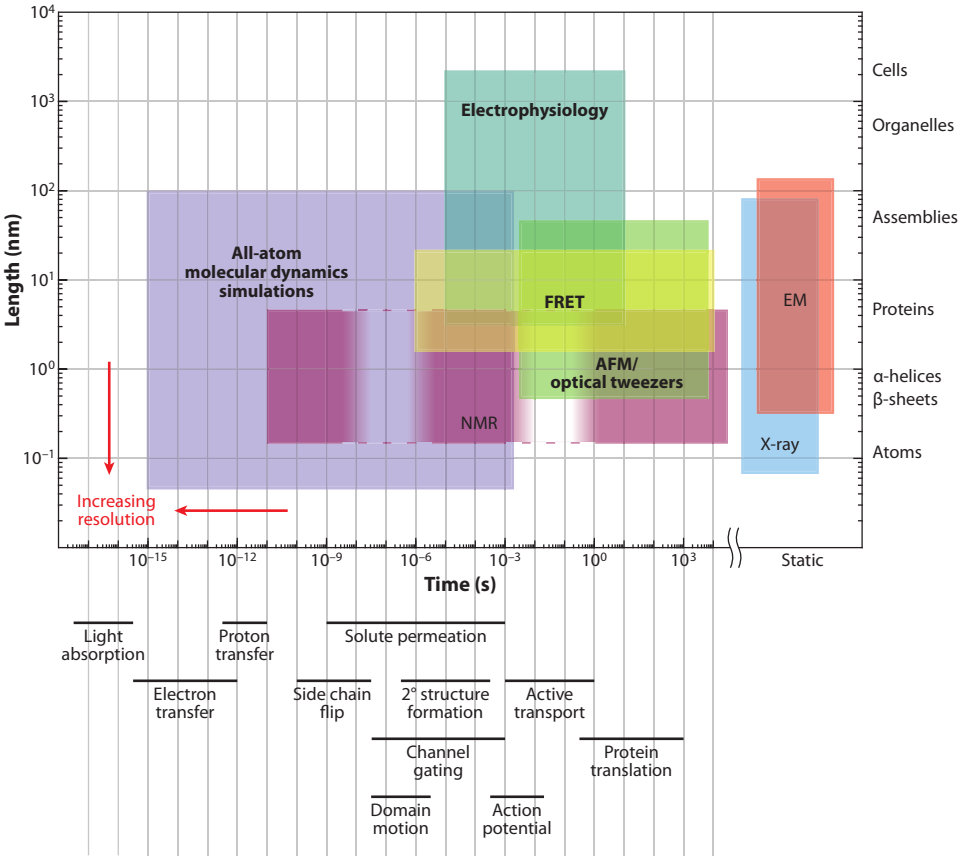


Figure 1.1: Spatiotemporal resolution of various biophysical techniques. The temporal (abscissa) and spatial (ordinate) resolutions of each technique are indicated by colored boxes. Techniques capable of yielding data on single molecules (as opposed to only on ensembles) are in boldface. NMR methods can probe a wide range of timescales, but they provide limited information on motion at certain intermediate timescales, as indicated by the lighter shading and dashed lines. The timescales of some fundamental molecular processes, as well as composite physiological processes, are indicated below the abscissa. The spatial resolution needed to resolve certain objects is shown at the right. Abbreviations: AFM, atomic force microscopy; EM, electron microscopy; FRET, Forster resonance energy transfer; NMR, nuclear magnetic resonance. Figure from [5], reproduced with permission.

An alternative to resolving these experimental limitations, in principle, is to model atomistic motions computationally, based on first principles physics. The computational expense, combined with the challenge of developing appropriate physical models, has limited both their length and their accuracy. A timeline

table (Table 1.1) contextualizes and summarizes the historical progress step by step in molecular dynamics, mainly in the biomolecules targeted in this thesis: proteins, lipids and logically combined systems. Historically, the timescales accessible to MD simulation have been shorter than those on which most biomolecular events of interest take place (see Table 1.1). This is the limit of the applicability of these simulations. Events such as protein folding, protein-drug binding, and major conformational changes essential to protein function typically take place on timescales of microseconds to milliseconds (Figure 1.1). MD simulations have generally been limited in practice to nanosecond timescales. Simulations of even a few microseconds required months on the most powerful supercomputers available, and longer simulations had never been performed. However, coarse-grained methods appear like a real option for performing such simulations as will be mentioned in the section 1.3.

It is important to highlight that the Bovine Pancreatic Trypsin Inhibitor (BPTI), a 58 amino acid protein has been used very often as the subject of choice in the early days of MD computer simulations. First because of its small size and second because it was one of the first proteins for which a high resolution crystal structure was available. Different papers might be honored as the first simulation of a water solvated protein, all of them using BPTI. This protein has been used as the reference protein in the historical progress of molecular dynamics shown in Table 1.1.

One of the main topics of molecular dynamics simulations has been the folding of proteins. Along the "short" and "fast" history of the molecular dynamics simulations of proteins, it is hard to define the first successful atomistic simulations of a protein in explicit solvent at physiological temperature by looking at the timeline of Table 1.1. This is due to many examples reported of protein unfolding, partial folding or folding of small peptides. To pinpoint to some examples, the work mentioned in Table 1.1 at 1975 was only marginally successful in folding, but was the first attempt to fold a protein using molecular dynamics.

On the other hand, it is interesting to address molecular dynamics advances in function of how the computational cost has been continously reduced, as is shown in Table 1.1. The report shown at 2010 in Table 1.1 indicated that long time-scales MD simulations could be achieved using a computer specifically designed for MD simulations [31], called *Anton*. Although this super-computer can generate long simulations it is not available for use to the general community and this is a disadvantage.

Table 1.1: Timeline of the most important results in the history of molecular dynamics simulations of both proteins and lipids.

1955	MD of an one-dimensional string of beads[6]
1957	MD of a two-dimensional system with which they called "spheres"[7]
1959	First extensive description of the MD method and algorithms[8]
1964	MD simulation of a system of 864 $Ar_{(l)}$ atoms[9]
1971	First MD simulation of water[10]
1975	First "*coarse grained*" MD simulation of BPTI protein: residues represented by only one/two interaction sites[11]
1976	Onset to the simulations of lipid bilayers using a two dimensional system of 240 dumbbell shaped "molecules"[12]
1977	First atomistic simulation of a BPTI protein contained four water molecules (called "united atom" model)[13]
1980	First atomistic simulation of a monolayer (not explicit water)[14]
1981	Water models: Simple Point Charge (**SPC**) and Transferable Inter-molecular Potential functionS (**TIPS**)[15, 16]
1982	First atomistic simulation of a bilayer (not explicit water)[17]
1984	BPTI in a truncated octahedron surrounded by 1467 water molecules for 20 ps[18]
1987	BPTI in 4785 water molecules for 42 ps in 1987[19]
1986	First study of a micelle in explicit water (1094 water molecules)[20]
1988	– BPTI solvated in 2607 water molecules for 210 ps[21] – First bilayer with explicit water (526 water molecules for 180 ps)[22]
1994	First study of an atomistic phospholipid (DPPC) bilayer[23, 24]
1995	Lipids and proteins: simulating bacteriorhodopsin in a lipid bilayer[25]
1998	– First attempt to fold a protein using molecular dynamics[26] – First 1 μs simulation of the peptide Villin Headpiece in water[27]
2001	First simulation of the self-assembly of a lipid bilayer (64 randomly oriented DPPC-molecules)[28]
2002	First folding study of a 23 residue protein (β-hairpin, turn and α-helix) in implicit solvent[29]
2010	First MD study reaching the millisecond for a single trajectory (23 residue WW-domain designed)[30]

1.2 Building a MD model

Components of a MD Simulation

The progress in MD simulations has faced different problems such as computational cost, size of proteins, number of particles or folding and unfolding phenomena with the development of new algorithms. As a consequence there has been

a growth of new, more sophisticated and complex software. However, the main components of a MD model remain unchanged.

Basically, to build a MD model it is necessary (combining software and force fields) to input *the molecular coordinates* and to choose *the force field* that describes the interactions between all atoms in the molecules. The software provides the *algorithms* to integrate the laws of motion and evolve the system through time to maintain *system properties* such as temperature and pressure. In this section a partial overview on this topic is described. For a complete discussion there are many reviews, however, those written by Gunsteren are particularly illustrative. [32, 33].

1.2.1 Force Fields and Interaction Functions

The term *"force field"* refers to the combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates. This set of parameters for interaction potentials takes a specific form which differs from one force field to another. The most important and most widely used atomistic force fields are **AMBER** (Assisted Model Building with Energy Refinement), **CHARMM** (Chemistry at HARvard Macromolecular Mechanics), **GROMOS** (GROningen MOlecular Simulation) and **OPLS** (Optimized Potentials for Liquid Simulations). A direct comparison between these force fields is given by Guvench et al.[34]. The simulation packages contain the algorithms that hand the laws of physics. These packages mostly appeared around the same time as the force fields and are often co-developed. The software also includes implementations of the most common algorithms in MD simulations together with programs to setup and analyze simulations. Well-known examples are GROMACS, GROMOS and NAMD (Not *just* Another Molecular Dynamics program)[34].

These force fields were built in the 1980s and different versions have been published since then. In spite of the many differences between them, they share some basic characteristics which can be subdivided into three parts:

(i) Non-bonded

Van der Waals interactions that can be described by the known Lennard-Jones (LJ) or Buckingham potentials, and **Electrostatic** interactions described by the Coulomb potential. The non-bonded interactions are computed on the basis of a neighbor list (a list of non-bonded atoms within a certain radius), in which exclusions are already removed.

(ii) Bonded

Covalent bond-stretching, angle-bending, improper dihedrals, and proper dihedrals. These are computed on the basis of fixed lists.

(iii) Others and Restraints

Position restraints, angle restraints, distance restraints, orientation restraints and dihedral restraints are all based on fixed lists.
The first two terms are inherent to the force field form and both can be represented by the well–known expressions:

$$E_{bond} = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi[1+\cos(n\chi-\sigma)] \quad (1.1)$$

and

$$E_{non\ bond} = \sum_{\substack{nonbond \\ pairs\ ij}} \left(\varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right) \quad (1.2)$$

where E_{bond} and $E_{nonbond}$ are the bonded and non-bonded interactions contributions to the total energy. The total energy is then:

$$E_{total} = E_{bond} + E_{nonbond} + E_{others} \quad (1.3)$$

Very often, the LJ term in the equation (1.2) is written as the functions of the pair of parameters A_{ij} and B_{ij} where $A_{ij} = \varepsilon_{ij} R_{min,ij}^{12}$ and $B_{ij} = 2\varepsilon_{ij} R_{min,ij}^{6}$. As an alternative form, the LJ potential may also be written as:

$$V_{LJ}(\mathbf{r}_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \quad (1.4)$$

Where $\sigma_{ij} = 2^{1/6} R_{min,ij}$ corresponds to the finite distance at which the potential V_{LJ} is zero. The interactions (parameters) matrix is built with the different values of the parameters ε_{ij} and σ_{ij} according to the combinations rules set up by the MD software or for the force field. These combination rules can be Lorentz–Berthelot rules.

Although the equations (1.1) and (1.2) contain the basic forms to describe the interactions types between particles, there are some differences or similarities in the different force fields. In the equations and the force fields there is no explicit term for hydrogen bonding: but the biologically important hydrogen bonds are handled by the combinations of the LJ and Coulomb terms.

1.2.2 The Algorithms

The following section describes the laws of physics and the algorithms handling them, ranging from Newton's laws of motion and the leap-frog integrator to schemes to handle electrostatics and chemical calculations.

(i) Integration Schemes

Newton's equation of motion can be integrated using a simple algorithm in small time steps Δt. Using a precedure with Taylor expansions we obtain equations which form the so-called leap-frog scheme. It is one, the most accurate, stable, and yet simple and efficient algorithms available for molecular dynamics simulations. There are several things worth mentioning about this algorithm: high number of force evaluations per time step, appropiate order of the algorithm and minimal memory storage and computational requirements.

(ii) Application of Constraints

Constraint methods are used for molecular bond lenghts and bond angles. They are used to save computing time. The length of the time step Δt in MD simulation is limited by highest frecuency (ν_{max}) motions occurring in the system. By freezing the generally uninteresting high-frequency internal vibrations, such as bond-length or possibly bond-angle vibrations, ν_{max}^{-1} is increased, which allows for a longer time step Δt.

(iii) Searching Neighbors

The bulk (approximately 90%) of the computer time required by a MD simulation is used for calculating the non-bonded interactions (equation 1.2), that is, for finding the nearest neighbor atoms and subsequently evaluating the Van der Waals and Coulomb interaction terms for the atom pairs obtained. Two schemes for performing this task as efficiently as possible have been proposed: *scanning all possible atom pairs* in an operation proportional to N^2, and *grid search techniques* whose operation is proportional to N.

1.2.3 Statistical Considerations

The Thermodynamic properties of a system containing N particles may be determined by solving the 3N Newtonian equations of motion for infinite sampling time. In practice, this is both impossible and unnecessary. It is impossible because only the deterministic equations for systems composed of a handful of particles could be solved even with modern computing resources. It is unnecessary because a wealth of information may be obtained about a particular system by

concentrating on its probabilistic configurations (field of study of *statistical mechanics*). It consists of taking into account systems at equilibrium by probability functions, and with some elementary assumptions, connecting the microstates of a system to macroscopic thermodynamic quantities (such as temperature, chemical potential and free energy). Molecular dynamics simulation, which solves the Newtonian equations of motion for each atom in a system, allow connections to be made between statistical mechanics and the thermodynamics of molecular recognition [35].

On the other hand, since the beginning of protein molecular dynamics simulation, the issues of sampling have been a major concern. This is subject to the ergodicity principle (usually called *ergodic hypothesis*), "*The properties of a system are the same when averaged over time as averaged over many systems*". A number of reports have tested the degree of ergodicity in a simulation, but there are still some discussions about how to demostrate that a simulation is not converged or that it may be converged. In some sense, a priori any molecular dynamics simulation has not fully sampled the phase space, because some events, such as spontaneous unfolding of a soluble protein under native conditions, could not be seen [36, 37].

The set of all states of the system is called an "*ensemble*", the mathematical space in which it is described is called "*phase-space*". Depending on which constraints are applied to the system different ensembles can be obtained. For example: a system of classical particles that is isolated from the outside world such that the macroscopic properties N (number of particles), V (the volume), and E (the total energy of the system) remain constant. Such a system is referred to as the **microcanonical ensemble** or the NVE ensemble (for constant Number, Volume, and Energy). This system can be in one of V accessible microstates, if one assumes that transitions to and from each microstate are permissible (again, the *ergodic hypothesis*). However, for our purposes, a more pertinent system is one that still has a constant N and V, but that is in thermal equilibrium with its environment and thus has a non-constant E but a constant temperature T. This is the **canonical ensemble**, or NVT (constant Number, Volume, and Temperature) ensemble. The canonical ensemble assumes a constant number of particles, volume and temperature; however, it is often desirable to consider a system in which the volume is allowed to fluctuate to maintain a constant pressure. This is referred to as the **isobaric-isothermal ensemble**, or NPT (constant Number, Pressure, and Temperature) ensemble [35, 38]. As a more general case, the system of interest may be allowed to exchange particles with its environment in the **Grand Canonical Ensemble**, where the chemical potential is considered a constant instead of the number of particles in the system. In general, the canonical ensemble provides a sufficient description for our purposes, making the grand canonical ensemble beyond the scope of this thesis.

As a final consideration in this section, the *ergodicity principle* for systems

in equilibrium, running one long simulation, will give the same result as running multiple short simulations. In this framework, one can do multiple simulations at the same time on different computers. However, this assumes that the separated systems are in equilibrium, which for some processes might not be feasible in short simulations. For these cases most of the software packages mentioned before have implemented parallelization schemes that allow a single simulation to run on multiple processors, by distributing the calculations or by changing the seed number for each simulation. In this way, longer simulations of larger systems can be simulated at the same time by using very large computers [39].

1.3 Coarse Graining

Coarse graining makes reference to the omission of irrelevant degrees of freedom of a system, in order to simplify the model. Strictly speaking every potential used for MD simulations is in some sense a coarse-grained potential. For example, in all-atom force fields the dynamics of electrons and nuclei are not explicitly considered (most force fields) and for example united-atom force fields (like GROMOS) do not explicitly consider carbon-bound hydrogen. However, in this section and in this thesis Coarse Graining (**CG**) will refer to those force fields that do not explicitly consider every heavy atom as a separate interaction site. In other words, the term coarse graining will mean the existence of a lower resolution in comparison to atomistic models. This does not mean that CG force fields are always derived from an atomistic force field.

As shown earlier in Table 1.1, in 1975 Levitt made a first approximation to molecular dynamics coarse-grained simulation using the BPTI protein represented as residues with one or two interaction sites. In 1976 Cotterill represented a lipid using simple dumbbell shaped objects. These examples show that CG force fields can have a wide range of different resolutions (the number of heavy atoms mapped to one CG site) and very different methods can be used to construct the potential energy function. A compendium of many different methods, force fields and applications is given in [40].

It is possible to distinguish three categories in the methods used to obtain a CG force field: *(a)* entirely based upon an atomistic force field, *(b)* entirely independent of any atomistic force field and *(c)* combinations of these two. The first category involves methods like inverse Boltzmann[41] and force matching[42, 43], where both bonded and non-bonded terms are directly derived from atomistic simulations. The second category includes force fields that have been purely parametrized to match macroscopic properties of the system to experimental quantities, like bilayer properties or protein packing data derived from the Protein Data Bank (PDB). Examples of this type of force field are described in [44, 45]. Methods in the third category often derive non-bonded interactions from

experimental quantities, like partitioning free energies or liquid properties, while they (partially) rely on atomistic simulations for bonded interactions. Examples of this last category are the force fields derived by DeVane and coworkers[46] and the Martini force field[47]. Thus, this thesis will focus on the Martini force field due to its simplicity to apply [48] and its versatility of use in different systems [28, 47, 49, 50].

Martini Force Field

The Martini CG force field applies a mapping of (on average) four heavy atoms to one CG interaction site. Four CG bead types have been defined: charged (Q), polar (P), non-polar (N) and apolar (C), which in turn are subdivided in four or five levels, giving a total of 18 bead types. For the interactions between these bead types, 10 different interactions levels are defined (O–IX) [47]. The interactions between these beads are Lennard-Jones potentials at ten different levels, corresponding to different values of the interaction parameter ϵ. The interaction between two beads has been parametrized by matching the free energies of vaporization, hydration and partitioning between water and apolar solvents for model compounds. These model compounds are chemical entities covering a wide range of molecules. Each of them are represented by one bead type and serve as building blocks for larger molecules. This modular setup and relatively small number of pre-parametrized bead types make it easy to build new molecules compatible with the Martini force field. The initial publication contained parameters for water, organic solvents, surfactants, lipids and cholesterol [47, 49]. In later publications parameters were added for more types of molecules, including proteins [48].

The loss of resolution in the Martini model compared to atomistic force fields brings along different limitations and challenges. Except for the obvious loss in structural detail, there are a few problems that especially are worth mentioning.

First, the grouping together of four heavy atoms reduces the entropy in a molecule. In order to obtain correct free energies, this is corrected for by adapting the enthalpic interactions [51]. As a result, the balance between entropy and enthalpy will be disturbed and separating these two contributions has to be done with the greatest care.

Second, a mapping of four heavy atoms to one CG interaction site, means four water molecules are grouped together in one bead. This works fine, unless one wants to study the behavior of single water molecules, for example water inside water channels in proteins or interfacial water at protein interfaces. Note that grouping four water molecules to one bead leads to a decrease in the number of interaction sites by a factor of 12 and is also responsible for the mass used for beads in Martini: 72 amu.

Third, coarse grain beads do not explicitly represent a possible polar nature

of the underlying molecular building block. The polar nature is mimicked by a slightly stronger interaction between polar beads and charged beads. In the case of a four water cluster the polarizability is also lost. For water this has been solved by introducing polarizable Martini water [52] and for polar amino acids a solution is introduced by Djurre et al [53].

Fourth, the secondary and tertiary structure of the proteins are not stable due to the lack of detail in the (backbone) interactions and the resulting lack of polarity mentioned above. For the secondary structure this has been solved by constraining α–helices and β–strands by angles, dihedrals and local elastic networks [48]. The tertiary structure can be kept stable by using a global elastic network, the so-called ElNeDyn approach, which connects backbone beads in different parts of the protein by long elastic bonds [54].

Compared to united atom models, the coarse grain mapping decreases the number of interaction sites by four, and for water by 12. The resulting total speed-up of a simulation, as compared to a united-atom force field is between a factor of 100 to 1000. There are three contributions to the speed-up. First, a reduced number of interaction sites means less calculations to be performed per time step. Notice that this goes roughly with the square of the reduction in the number of particles, since not only the number of particles for which the interactions have to be calculated is reduced, but also the number of interactions to other particles. For Martini, this would be on average $4^2 = 16$, although for water it is for more ($12^2 = 144$). Second, coarse graining leads to a smoother energy landscape due to the omission of fast modes in the molecules. The smoother energy landscape allows for larger time steps without making large integration errors. For Martini, typically a time step between 20 and 40 fs can be applied [55], 4 to 40 times longer than those used in atomistic simulations. Last, the smoother energy surface is also responsible for a faster progress in dynamic processes, for example diffusion. In others words, in the same simulation time a system will sample a larger part of phase space.

References

[1] Levitt M. *Nat. Struc. Mol. Biol.*, 8:392–393, 2001.

[2] Karplus M. *Biopolymers*, 68:350–358, 2003.

[3] Schlick T, Collepardo-Guevara R, Halvorsen LA, Jung S, and Xiao X. *Quart. Rev. Biophys.*, 44:191–228, 2011.

[4] Vendruscolo M and Dobson CM. *Curr. Biol.*, 21:R68–R70, 2011.

[5] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. *Annual Review of Biophysics*, 41(1):429–452, 2012.

[6] Fermi E, Pasta J, and Ulam S. *Los Alamos Report*, LA-1940, 1955.

[7] Alder BJ and WainWright TE. *J. Chem. Phys.*, 27:1208–1211, 1957.

[8] Alder BJ and WainWright TE. *J. Chem. Phys.*, 31:1–9, 1959.

[9] Rahman A. *Phys. Rev.*, 136:405–411, 1964.

[10] Rahman A and Stillinger FH. *J. Chem. Phys.*, 55:3336, 1971.

[11] Levitt M and Warshel A. *Nature*, 253:694–698, 1975.

[12] Cotterill RMJ. *BBA-Biomembranes*, 433:264–270, 1976.

[13] McCammon JA, Gelin BR, and Karplus M. *Nature*, 267:585–590, 1977.

[14] Kox AJ, Michels J, and Wiegel F. *Nature*, 287:317–319, 1980.

[15] Berendsen H, Postma J, van Gunsteren W, and Hermans J. *Intermolecular forces*, chapter Interaction models for water in realtion of protein hydration, pages 331–342. Reidel, Dordrecht, 1981.

[16] Jorgensen WL and William L. *Journal of the American Chemical Society*, 103(2):335–340, 1981.

[17] V der Ploeg P and Berendsen HJC. *J. Chem. Phys.*, 49:233–248, 1982.

[18] Van Gunsteren WF and Berendsen HJC. *J. Mol. Biol*, 176:559–564, 1984.

[19] Wong CF and McCammon JA. *Isr. J.Chem*, 27:211–215, 1987.

[20] Jönsson B, Edholm O, and Teleman O. *J. Chem. Phys.*, 85:1–14, 1986.

[21] Levitt M and Sharon R. *Proc. Nati. Acad. Sci. USA*, 85:7557, 1988.

[22] Egberts E and Berendsen HJC. *J. Chem. Phys.*, 89:3718, 1988.

[23] Egberts E, Marrink SJ, and Berendsen HJC. *Eur. Biophys. J.*, 22:423–436, 1994.

[24] Marrink SJ, Berkowitz M, and Berendsen HJC. *Langmuir*, 9:3122–3131, 1993.

[25] Edholm O, Berger O, and Jähnig F. *J. Mol. Biol.*, 250:94–111, 1995.

[26] Daura X, Jaun B, Seebach D, van Gunsteren WF, and Mark AE. *J. Mol. Biol.*, 280:925–932, 1998.

[27] Duan Y. *Science*, 282:740–744, 1998.

[28] Marrink SJ, Lindahl E, Edholm O, and Mark AE. *J. Am. Chem. Soc.*, 123:8638–8639, 2001.

[29] Snow CD, Nguyen H, Pande VS, and Gruebele M. *Nature*, 420:102–106, 2002.

[30] Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, and Shan Y. *Science*, 330:341, 2010.

[31] Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, Young C, Deneroff MM, Batson B, and Bowers KJ. *P. C. High Perf. Comp. Netw. Stor. Anal.*, page 39, 2009.

[32] Wilfred F. van Gunsteren and Herman J. C. Berendsen. *Angewandte Chemie International Edition in English*, 29(9):992–1023, 1990.

[33] Wilfred F. van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glättli, Philippe H. Hünenberger, Mika A. Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt, and Haibo B. Yu. *Angewandte Chemie International Edition*, 45(25):4064–4092, 2006.

[34] Guvench O and MacKerell AD. *Method. Mol. Biol.*, 443:63–88, 2008.

[35] Jeff Wereszczynski and J. Andrew McCammon. *Quarterly Reviews of Biophysics*, 45:1–25, 2 2012.

[36] RD. Mountain and D. Thirumalai. *The Journal of Physical Chemistry*, 93(19):6975–6979, 1989.

[37] Alan Grossfield, Scott E. Feller, and Michael C. Pitman. *Proteins: Structure, Function, and Bioinformatics*, 67(1):31–40, 2007.

[38] Berendsen HJC. *Simulating the physical world: hierarchical modeling from quantum mechanics to fluid dynamics.* Cambridge University Press, 2007.

[39] Hess B, Kutzner C, van der Spoel D, and Lindahl E. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[40] Voth GA. *Coarse-Graining of Condensed Phase and Biomolecular Systems.* CRC Press, 2008.

[41] Soper A. *Chem. phys.*, 202:295–306, 1996.

[42] Izvekov S, Parrinello M, Bernham CJ, and Voth GA. *J. Chem. Phys.*, 120:1–19, 2004.

[43] Noid WG, Chu J, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, and Andersen HC. *J. Chem. Phys.*, 128:244114, 2008.

[44] Cooke IR, Kremer K, and Deserno M. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, 72:011506, 2005.

[45] Miyazawa S and Jernigan RL. *J. Mol. Biol.*, 256:623–644, 1996.

[46] Shinoda W, DeVane R, and Klein ML. *Mol. Simulat.*, 33:27–36, 2007.

[47] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, and de Vries AH. *Journal of Physical Chemistry B*, 111(27):7812, 2007.

[48] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, and Marrink SJ. *Journal of Chemical Theory and Computation*, 4(5):819, 2008.

[49] Marrink SJ, de Vries AH, and Mark AE. *Journal of Physical Chemistry B*, 108(2):750, 2004.

[50] Marrink SJ and Tieleman DP. *Chem. Soc. Rev.*, 42(16):6801, 2013.

[51] Riccardo Baron, Alex H. de Vries, Philippe H. Hnenberger, and Wilfred F. van Gunsteren. *The Journal of Physical Chemistry B*, 110(31):15602–15614, 2006.

[52] Semen O. Yesylevskyy, Lars V. Schfer, Durba Sengupta, and Siewert J. Marrink. *PLoS Comput Biol*, 6:e1000810, 06 2010.

[53] Djurre H. de Jong, Xavier Periole, and Siewert J. Marrink. *Journal of Chemical Theory and Computation*, 8(3):1003–1014, 2012.

[54] Periole X, Cavalli M, Marrink S-J, and Ceruso MA. *Journal of Chemical Theory and Computation*, 5(9):2531, 2009.

[55] Siewert J. Marrink, Xavier Periole, D. Peter Tieleman, and Alex H. de Vries. *Phys. Chem. Chem. Phys.*, 12:2254–2256, 2010.

Chapter 2

Epidermal Growth Factor Receptors

Epidermal Growth Factor Receptors (**EGFR**) consists of four different members: EGFR/ErbB1, ErbB2, ErbB3 and ErbB4. They are associated with a number of biological processes and are becoming increasingly recognized as important therapeutic targets against cancer. Additionally, they are also proteins with a big size and therefore a study by long-timescale molecular dynamics simulation is hard to achieve. For that reason only their domains (extracellular, transmembrane and intracelullar) are considered by separated parts to perform any molecular modeling. On the ohter hand, the dynamic nature of these proteins is very related to their biological function. Molecular dynamic simulation appears as a well-established method for modeling such motions because it provides a characterization which is difficult to access experimentally. Taking into account both the size limitation and the molecular dynamics advantages, the protein–protein and protein–membrane interactions regarding the EGFR receptors suppose an interesting challenge that has been explored on this thesis.

2.1 Structure and Function

Epidermal Growth Factor Receptors (EGFR) control multiple cellular processes, including cellular proliferation, survival differentiation and migration [2]. The EGFR family consists of four different members: EGFR or ErbB1, ErbB2, ErbB3 and ErbB4. These receptors are formed by three differentiated domains: an extracellular ligand-binding domain or ectodomain (**ECD**) containing four subdomains (I-IV), a single transmembrane α-helix (**TM**) spanning the cellular membrane and intracellular juxtamembrane (**JM**), a tyrosine-kinase (**TKD**) and autophos-

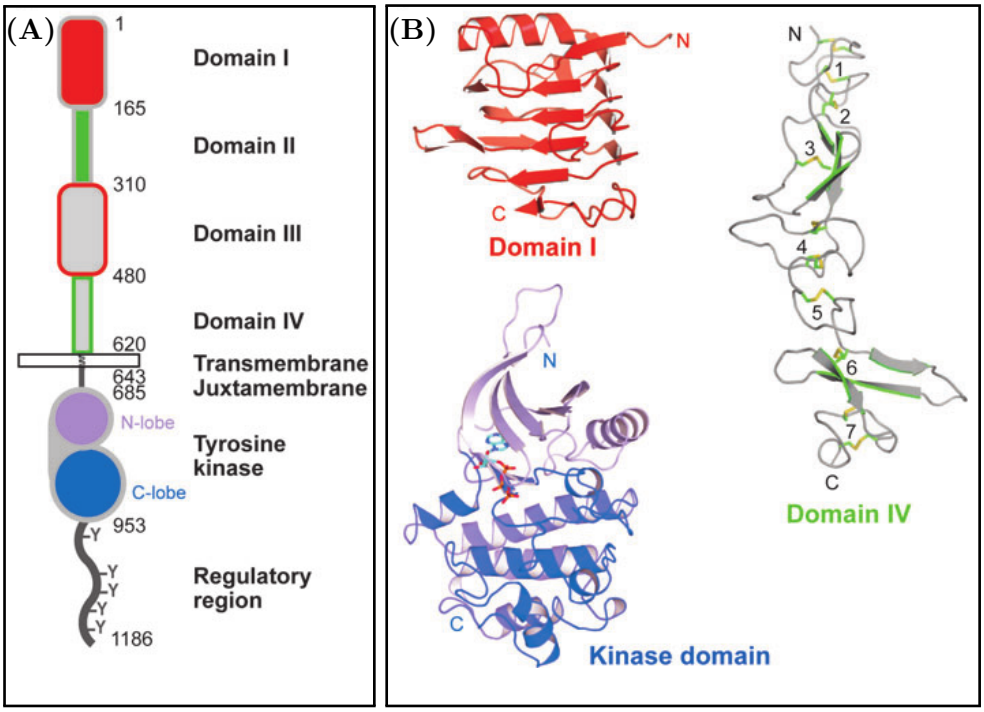


Figure 2.1: The domains of EGFR. (A) The extracellular region comprises four domains: I–IV, sometimes referred to as L1, CR1, L2, and CR2 or L1, S1, L2, and S2. Domains I (red) and III (gray with red outline) share about 37% sequence identity, while domains II (green) and IV (gray with green outline) are cystine rich. The N-lobe of the kinase domain is in lavender and the C-lobe is in blue. Amino acid numbers are noted for each domain boundary. (B) Representative ribbon diagrams of the domains of EGFR. Domains I and III adopt a β-helix fold; here domain I from PDB ID 1YY9 is shown. Domains II and IV adopt extended structures comprising a series of disulfide-bonded modules. Domain IV from PDB ID 1YY9 is shown with the disulfides in stick representation and the disulfide-bonded modules numbered. There are two types of disulfide-bonded module. One has a single disulfide bond and the intervening loops adopt a bow-like arrangement (modules 2, 3, 5, and 6). The second type has two disulfide bonds with consecutive cysteines linked in the pattern Cys1-Cys3 and Cys2-Cys4 (modules 1, 4, and 7). The inactive kinase is shown (PDB ID 2GS7) with the ATP analogue (AMP-PNP) in stick representation. Figure from [1], reproduced with permission.

phorylation domains. A schematic representation of these domains is shown in Figure 2.1. On the other hand, it is well established that epidermal growth factors (EGF) bind to the ECD domain, promoting EGFR dimerization and increasing the tyrosine-kinase activity of its intracellular domain (See Figure 2.1) [1, 3].

The orientation of the four ECD subdomains has been revealed by recent crystallographic studies (see Figure 2.2 for illustration). On the basis of these studies, both *tethered* (ErbB3 [4], ErbB1/EGF [5], ErbB1/Cetuximab [6], ErbB4

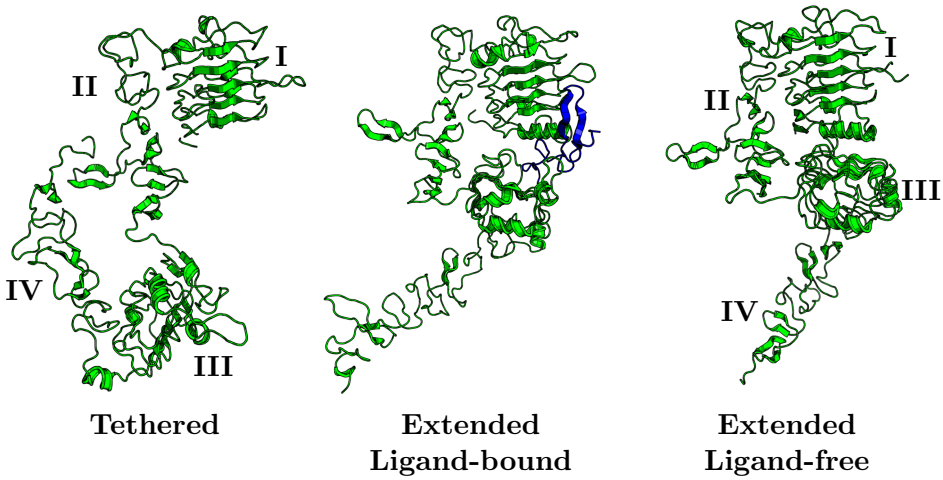


Figure 2.2: Possible Conformational Structures of Ectodomain. Roman numbers indicate the different subdomains. In the **tethered** structure the subdomain II are bound to subdomain IV (Structure from PDC code: 1M6B and common for ErbB1, ErbB3 and ErbB4). The **extended ligand-bound** structure exhibits the subdomains I binds to subdomain III mediated by ligand in blue (Structure from PDB code: 3NJP and common for ErbB1, ErbB3 and ErbB4). The **extended ligand-free** structure shows the subdomains I binds to III with no ligand (Structure from PDB code: 1N8Z and only for ErbB2)

[7]) and *extended* (EGF-bound [8] and TGF-α bound [9] truncated ErbB1 dimers, Trastuzumab [10] and pertuzumab [11] ErbB2 complexes and full EGF-bound ErbB1 dimer [12]) conformations have been proposed for ECD-EGFR receptors (Figure 2.2). The tethered or autoinhibited conformation ties subdomains II and IV, forming intramolecular interactions. Accordingly, the dimerization arm in subdomain II is buried by subdomain IV, impairing the formation of EGFR dimers. On the contrary, interactions between subdomains II and IV are broken in the active untethered or extended conformation, provoking a release of the dimerization arm, which can now dimerize with other EGFR monomers.

The ECD of ErbB1, ErbB3, ErbB4 contains a specific region where the EGF ligand is bound (Figure 2.2). This binding is thought to promote intracellular signaling cascades that regulate cellular growth and proliferation. Ligand binding promotes signaling activation by homo or hetero-dimerization in the membrane surface, changing the receptor conformation from a tethered structure to an extended dimerization-exposed arm conformation [13]. Other authors, however, claim that the ligand can alter the equilibrium between previously formed active/inactive dimers [13, 14]. Unlike other members of the family, the ErbB2

receptor can adopt a unique conformation (extended conformation) resembling the ligand-activated state, which is able to dimerize even in the absence of ligand [15]. These receptors are engaged in the regulation of many processes such as cell proliferation, differentiation and apoptosis. The abnormal regulation of these receptors generates a number of human diseases, such as cancer [2, 16].

The biological activity of these receptors is known to depend on the formation of homo or heterodimers in such a way that the intracellular kinase domains in close contact can interact, initiating a signaling cascade that ends with cellular proliferation [17]. The complex formed by ErbB2 and ErbB3 has been recognized as being of pivotal importance in cancer development and evolution [18].

2.2 Structure as Revealed by MD Simulations

In 1994, Garnier et al [19] performed the firt MD simulation reported about EGFR family. It was the transmembrane domain of ErbB2 (27 residues). In this study a 160-ps simulation was performed in vacuum using GROMOS force field and the influence of the mutation Val(659)→Glu(659) was examined. Later, the same team published at 1996 another 160-ps simulation with two transmembrane peptides of ErbB2 in vacuum [20]. These studies reported that a single amino acid replacement generates a propagation of local conformational changes along the transmembrane segment which is important in the signal transduction mechanisms of transmembrane receptors. Garnier at el [21] reported at 1997 the first trasnmembrane dimer of ErbB2. They performed a MD simulation by 575 ps without water. Two years later, Duneau at el [22] reported the first work of an ErbB receptor within an explicit membrane environment with a trasnmembrane ErbB2. It was a MD simulation by 500 ps using a fully solvated dilauroylphosphatidyl-ethanolamine bilayer (DLPE). They reported that one membrane simulation shows that the initial α-helix undergoes a local π-helix conversion in the peptide part embedded in the membrane core similar to that found in simulation vacuum.

According to the historical progress on computational resources (see Chapter 1) both protein size and time scale have increased along the years. That is the reason why the first studies on EGFR family were on the smaller domain: transmembrane region, particularly on ErbB2. Consequently, just in 1999, the first approximation on the tyrosine kinase domain by molecular dynamics was done. They performed a 1-ns simulation of ErbB1 with water [23]. However, a more extensive study on TK domain was performed by Suenaga [24] in 2003. A simulation by 1 ns with 98 residues belonging to SH2 domain (see Figure 2.1), fully solvated.

Finally, it was in 2005 when the first report about ectodomain was done. Luo et al [25] performed MD simulations on homology structures for ErbB3 and

ErbB4. Those structures were built using as template the X-ray structure of 2:2 EGF:EGFR complex (PDB ID: 1IVO). A total of 1 ns of time simulation was performed in this study.

On the other hand, the first microsecond MD simulations, was reported in 2012. Du et al. [26] performed microsecond MD simulations of ErbB4 tethered with its endogenous ligand neuregulin–1 β (NRG–1 β). While the conformational transition of the ECR-ErbB4/NRG–1β complex from a tethered inactive conformation to an extended active-like form is observed clearly in the simulation, the conformational change of ECD-ErbB4 is not. Therefore, it could be proposed that ligand binding is indeed the active driving force for the conformational transition and further dimerization to occur. These authors constructed an energy landscape for the conformational transition of ECD-ErbB4/NRG–1β complex and reported that the energy barrier for the tether opening has a value of 2.7 kcal/mol, which is in agreement to the experimental value (1-2 kcal/mol) reported for ErbB1.

In adition of those mentioned above, relevant studies on tyrosine kinase domain have been performed by Telesco team [27–30]. They have revelead by MD simulations transitions between inactive and active conformations in almost every EGFR family receptors. Time simulations of 10 ns were used in their studies.

Finally, the study of the structure of EGFR receptors has been performed to know the dynamic behavior. Thus, due to the X-ray or NMR structure absence, the molecular dynamics has been used to predict structure, intra and/or inter molecular interactions and conformational flexibility. Due to the experimental difficult to get a complete structure of the EGFR receptor, only separated domains have been studied by structural techniques (X-ray or NMR) and at the same time by atomistic molecular dynamics. However, in 2013, another related system studied by atomistic molecular dynamics simulation is the near half million atom complex simulated over tens of microseconds by Shaw's group on the Anton supercomputer, a "hardware" specifically designed to accommodate molecular dynamics algorithms (see section 1.1). This system consisted of a complete EGFR dimer including the lipid bilayer [31]. This work reported that, in ligand-bound dimers, the ectodomains take conformations favoring dimerization of the transmembrane helices, dimerization of the juxtamembrane segments, and formation of asymmetric kinase dimers. In ligand-free dimers, the ectodomain does not favor the N-termini interaction in the transmembrane domain, then, promote the juxtamembrane segment dissociation and formation of symmetric kinase dimers.

Our molecular dynamics simulations of membrane-embedded EGFR suggest that, in ligand-bound dimers, the extracellular domains assume conformations favoring dimerization of the transmembrane helices near their N termini, dimerization of the juxtamembrane segments, and formation of asymmetric (active) kinase dimers. In ligand-free dimers, by holding apart the N termini of the transmembrane helices, the extracellular domains instead favor C-terminal dimerization

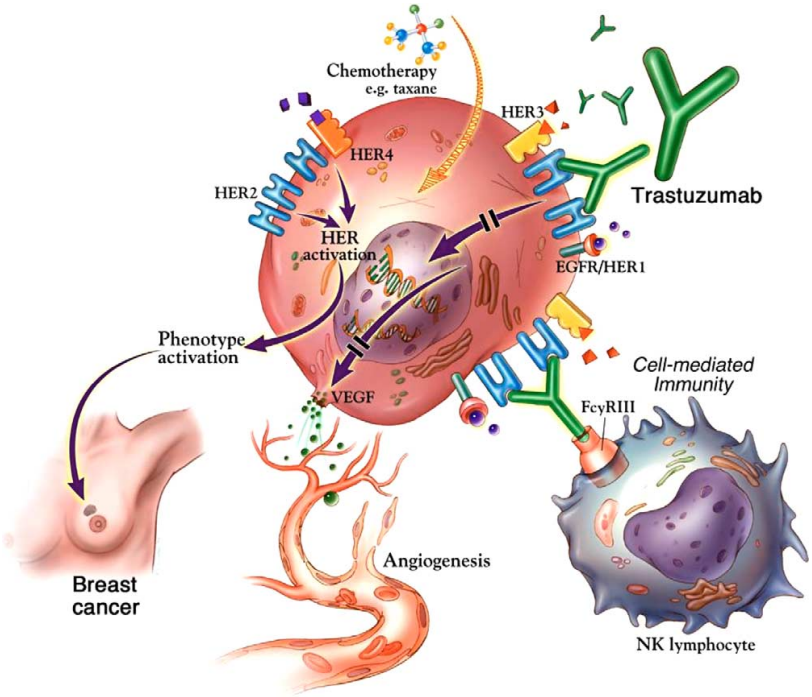


Figure 2.3: Proposed mechanisms of action of Herceptin (Trastuzumab). The illustrated mechanisms are described in detail in the text. Figure from [33], reproduced with permission.

of the transmembrane helices, juxtamembrane segment dissociation and membrane burial, and formation of symmetric (inactive) kinase dimers. Electrostatic interactions of EGFRs intracellular module with the membrane are critical in maintaining this coupling.

2.3 ErbB2 and Trastuzumab Antibody

The overexpression of ErbB2 leads to EGFR receptor activation in tissue culture, while overexpression of other EGFR receptors is not active unless a ligand is added [32].

Trastuzumab (Tzb), currently constitutes a part of the immunotherapy treatment of advanced breast cancers, i.e., those with extensive metastasis, and in general, solid tumors overexpressing ErbB2. Different clinical trials confirm the efficiency of the antibody as an anticancer treatment. The monoclonal antibody Trastuzumab can bind to ECD-ErbB2 domain [10] (see Figure 2.3). The fragment antigen binding (Fab) of Trastuzumab that binds to subdomain IV of

the ECD-ErbB2 categorizes this site as a possible target for anticancer therapies. Additionally, some authors claim, based on these structural studies, that Trastuzumab is not effective in blocking dimerization of ErbB2 with ligand activated EGFR or ErbB3 [34, 35]. However it has been reported that ligand independent ErbB2/ErbB3 complex is disrupted by Trastuzumab [35].

The monoclonal antibody Trastuzumab was designed to target an extracellular epitope of the ErbB2 receptor [36–38], being the first specific anti-ErbB2 treatment approved by the Food and Drug Administration (FDA). However, the molecular mechanism of action of Trastuzumab is not at all understood. Different effects have been reported regarding its anticancer activity [33, 39, 40]. For example, Trastuzumab has been identified as an inhibitor of the downstream signaling cascade for cellular proliferation by its action through a number of different mechanisms (see Figure 2.3).

Those mechanisms are still incompletely defined, however, several molecular and cellular effects have been observed in experimental *in vitro* and *in vivo* models, which can be the following:

(i) Internalization and degradation of ErbB2. Disrupts receptor dimerization; disrupts downstream signaling pathways

(ii) G1 arrest and reduced proliferation. The G1 phase, or Growth 1/Gap 1 phase, is the first of four phases of the cell cycle that takes place in eukaryotic cell division. Thus, Induces $p27^{kip1}$-cdk2[1] complex formation; induces $p27^{kip1}$ levels.[2]

(iii) Apoptosis. Inhibits Akt[3] activity

(iv) Suppresses angiogenesis. Reduces tumor vasculature in vivo; reduces expression of pro-angiogenic VEGF[4], TGF-α, Ang-1, PAI-1; induces anti-angiogenic TSP-1

(v) Immune-mediated responses. ADCC[5]; stimulates natural killer cells

(vi) Inhibits ECD-ErbB2 proteolysis

(vii) Inhibits DNA repair

The ErbB2 epitope targeted by Trastuzumab is located in subdomain IV of the ECD. The tight interaction of the monoclonal antibody Fab with subdomain

[1] **cdk**, cyclin-dependent kinase

[2] $p27^{kip1}$ is a cell-cycle regulatory protein that interacts with cyclin-CDK2 and -CDK4, inhibiting cell cycle progression at G1

[3] **Akt**, alpha serine/threonine-protein kinase

[4] **VEGF**, vascular endothelial growth factor

[5] **ADCC**, antibody-dependent cellular cytotoxicity

IV of the ErbB2 was elucidated by crystallography [10]. Additionally, molecular dynamics simulations performed at the atomistic scale on the complex have shown an additional interaction between the antibody constant domain and the dimerization arm located in subdomain II of the receptor [41, 42]. The authors argued that the additional interaction does not appear in the crystallographic structures due to strong crystal packing conditions which are not present in solution (see Chapter 4).

Considering some computational works concerning ECDs, Fuentes et. al. [42] have carried out several molecular dynamics studies on ErbB2 along with two different antibodies, Trastuzumab (Tzb) and pertuzumab. They estimated the binding free energy for several receptor/antibody complexes. They reported that Trastuzumab has a higher affinity for apo ErbB2 than pertuzumab. Furthermore, the epitope for Trastuzumab is domain IV whereas pertuzumab is bound to the dimerization arm located in domain II. Subsequently, they found an increase in affinity when both antibodies are bound to the receptor.

On the other hand, the Pertuzumab antibody, which binds to subdomain II (dimerization arm), has shown to be effective to disrupt the ErbB2/ErbB3 ligand activated complex but ineffective for the independent ligand specie. In the light of these findings, it has been hypothesized that the ligand independent interaction between ErbB2/ErbB3 is different from the ligand-induced dimerization. Thus, the subdomain II – sudomain II interfaces may not mediate in the ligand-independent complex [35]. In contrast, very recent single-molecule force spectroscopy studies suggest a mechanism of blocking of the heterodimerization of ErbB2/Trastuzumab and ErbB3 receptors even in presence of the heregulin (HRG) ligand [43].

Additionally, Bagossi et al. [44] reported the first 2:2 ErbB2/Trastuzumab complex from homology based on the X-ray or nuclear magnetic resonance structures of extracellular, transmembrane, and intracellular domains. They predicted some favourable dimerization interactions for the extracellular, transmembrane, and protein kinase domains in the model of a nearly full-length dimer of ErbB2, which may act in a coordinated fashion in ErbB2 homodimerization. The Bagossi model was used in this thesis as the initial model to study the complete ErbB2 system.

Aim of this Thesis

According to the background previously mentioned, the following aims have been proposed to performed an extensive approximation between EGFR family receptor and the molecular dynamic simulations:

(i) Due to absences of the X-ray or NMR structure absences of ECD-ErbB3 and ECD-ErbB4 in their active forms, an homology and a subsequent molecular

dynamics refinement was targetted.

(ii) There is only X-ray structures for ECD-ErbB1 homodimers, therefore, to study one of most relevant dimers, ECD-ErbB2/ECD-ErbB3, by homology and molecular dynamcis refinement was proposed.

(iii) Study the conformational flexibility of the ECD-ErbB2/Trastuzumab complex by molecular dynamics at an atomic resolution.

(iv) Validate the Martini coarse-grained force field on the protein-protein interactions with the atomistic molecular dynamics results on the ECD-ErbB2/Trastuzumab complex.

(v) Perform coarse-grained molecular dynamics on the full-length ErbB2 receptor and learn about the protein-protein and protein-membrane interacions regarding the monomer, homodimer, antibody-monomer and antibody - homodimer systems.

As was mentioned along this Chapter, the size limitation of the EGFR receptors and the molecular dynamics advantages offer a very fascinating challenge to propose and explain the intra-protein and protein–protein interactions governing the EGFR receptors in their active conformation.

A conformation stabilized by molecular dynamics for each receptor described in the section 2.2 is proposed in Chapter 3. Additionally, their dynamic behavior in water is studied in an attempt to explain some results that are experimentally difficult to obtain.

Conformational flexibility of the ErbB2 ectodomain and Trastuzumab antibody complex is studied in Chapter 4. In this Chapter an exhaustive structural analysis as revealed by atomistic molecular dynamcis is performed.

This interesting complex is used as the subject to validate the proper parameters that should be used in molecular dynamics at coarse-grained resolution scale. This is performed as a comparative work between atomistic and coarse-graining in Chapter 5.

The comparison between these two scale resolution was used to perform an exploration of the dynamics and interaction of a full ErbB2 receptor and Trastuzumab-Fab antibody in a lipid bilayer model using Martini coarse-grained force field. These results are reported in Chapter 6. A detailed description of the type of interactions governing the homodimerization and antibody complexation phenomena and the role that the membrane plays on that is presented.

References

[1] Kathryn M. Ferguson. *Annual Review of Biophysics*, 37(1):353–373, 2008.

[2] Wieduwilt M and Moasser M. *Cell. Mol. Life Sci.*, 65(10):1566–1584, 2008.

[3] Jessica P. Dawson, Zimei Bu, and Mark A. Lemmon. *Structure*, 15(8):942 – 954, 2007.

[4] Cho H-S and Leahy DJ. *Science*, 297(5585):1330–1333, 2002.

[5] Ferguson KM, Berger MB, Mendrola JM, Cho H-S, Leahy DJ, and Lemmon MA. *Mol. Cell.*, 11(2), 2003.

[6] Li S, Schmitz KR, Jeffrey PD, Wiltzius JJW, Kussie P, and Ferguson KM. *Cancer Cell*, 7(4):301–311, 2005.

[7] Bouyain S, Longo PA, Li S, Ferguson KM, and Leahy DJ. *Proc. Natl. Acad. Sci. USA*, 102(42):15024–15029, 2005.

[8] Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim J-H, Saito K, Sakamoto A, Inoue M, Shirouzu M, and Yokoyama S. *Cell*, 110(6):775–787, 2002.

[9] Garrett TPJ, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Zhu H-J, Walker F, Frenkel MJ, Hoyne PA, Jorissen RN, Nice EC, Burgess AW, and Ward CW. *Cell*, 110(6):763–773, 2002.

[10] Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, and Leahy DJ. *Nature*, 421(6924):756–760, 2003.

[11] Franklin MC, Carey KD, Vajdos FF, Leahy DJ, de Vos AM, and Sliwkowski MX. *Cancer Cell*, 5:317–328, 2004.

[12] Lu C, Mi n Z, Grey MJ, Zhu J, Graef E, Yokoyama S, and Springer TA. *Mol. Cell Biol.*, 30(22):5432–5443, 2010.

[13] Lemmon MA. *Exp. Cell. Res.*, 315(4):638–648, 2009.

[14] Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, and Mellman I. *Nature*, 464(7289):783–787, 2010.

[15] Brennan PJ, Kumagai T, Berezov A, Murali R, and Greene MI. *Oncogene*, 19(53):6093–6101, 2000.

[16] Yarden Y and Sliwkowski MX. *Nat. Rev. Mol. Cell. Biol.*, 2(2):127–137, 2001.

[17] Hynes NE and Lane HA. *Nat. Rev. Cancer.*, 5(5):341–354, 2005.

[18] Baselga J and Swain SM. *Nat. Rev. Cancer*, 9(7):463–475, 2009.

[19] Norbert Garnier, Daniel Genest, Eric Hebert, and Monique Genest. *Journal of Biomolecular Structure and Dynamics*, 11(5):983–1002, 1994.

[20] Norbert Garnier, Daniel Genest, and Monique Genest. *Biophysical Chemistry*, 58(3):225–237, 1996.

[21] Norbert Garnier, Daniel Genest, Jean Pierre Duneau, and Monique Genest. *Biopolymers*, 42(2):157–168, 1997.

[22] Jean-Pierre Duneau, Serge Crouzy, Norbert Garnier, Yves Chapron, and Monique Genest. *Biophysical Chemistry*, 76(1):35 – 53, 1999.

[23] Günther H. Peters, Thomas M. Frimurer, Jannik N. Andersen, and Ole H. Olsen. *Biophysical Journal*, 77(1):505 – 515, 1999.

[24] Atsushi Suenaga, Mio Ichikawa, Mariko Hatakeyama, Xiaomei Yu, Noriyuki Futatsugi, Tetsu Narumi, Kazuhiko Fukui, Takaho Terada, Makoto Taiji, Mikako Shirouzu, Shigeyuki Yokoyama, and Akihiko Konagaya. *Biochemistry*, 42(18):5195–5200, 2003.

[25] Luo C, Xu L, Zheng S, Luo X, Shen J, Jiang H, Liu X, and Zhou M. *Proteins*, 59(4):742–756, 2005.

[26] Du Y, Yang H, Xu Y, Cang X, Luo C, Mao Y, Wang Y, Qin G, Luo X, and Jiang H. *J. Am. Chem. Soc.*, 134(15):6720–6731, 2012.

[27] Telesco SE and Radhakrishnan R. *Biophys. J.*, 96(6):2321–2334, 2009.

[28] Shi F, Telesco SE, Liu Y, Radhakrishnan R, and Lemmon MA. *Proc. Natl. Acad. Sci. USA*, 107(17):7692, 2010.

[29] Shih AJ, Telesco SE, Cho SH, Lemmon MA, and Radhakrishnan R. *Biochem. J.*, 436:241–251, 2011.

[30] Telesco SE, Shih AJ, Jia F, and Radhakrishnan R. *Mol. Biosyst.*, 7:2066–2080, 2011.

[31] Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, Kuriyan J, and Shaw DE. *Cell*, 152(3):557, 2013.

[32] DiFiore PP, Pierce JH, Segatto O, King CR, and Aaronson SA. *Science*, 237:78–182, 1987.

[33] Rita Nahta and Francisco J. Esteva. *Cancer Letters*, 232(2):123 – 138, 2006.

[34] Agus DB, Akita RW, , Fox WD, Lewis GD, Higgins B, Pisacane PI, Lofgren JA, Tindell C, Evans DP, Maiese K, Scher HI, and Sliwkowski MX. *Cancer Cell*, 2:127–137, 2002.

[35] Junttila TT, Akita RW, Parsons K, Fields C, Phillips GDL, Friedman LS, Sampath D, and Sliwkowski MX. *Cancer Cell*, 15:429–440, 2009.

[36] Baselga J, Tripathy D, Mendelsohn J, Baughman S, Benz CC, Dantis L, Sklarin NT, Seidman AD, Hudis CA, Moore J, Rosen PP, Twaddell T, Henderson IC, and Norton L. *J. Clin. Oncol.*, 14(3):737, 1996.

[37] Carter P, Presta L, Gorman CM, Ridgway JBB, Henner D, Wong WLT, Rowland AM, Kotts C, Carver ME, and Shepard HM. *Proc. Natl. Acad. Sci. USA*, 89(10):4285, 1992.

[38] Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, and Press M. *J. Clin. Oncol.*, 20(3):719, 2002.

[39] Nahta R and Esteva FJ. *Oncogene*, 26(25):3637, 2007.

[40] Nahta R. *Current Medicinal Chemistry*, 19(7):1065, 2012.

[41] JF Franco-Gonzalez, J Ramos, VL Cruz, and J Martínez-Salazar. *J. Mol. Model.*, 19(3):1227–1236, 2013.

[42] Fuentes G, Scaltriti M, Baselga J, and Verma C. *Breast Cancer Res.*, 13(3):R54, 2011.

[43] Shi X, Xu L, Yu J, and Fang X. *Exp. Cell Res*, 315:2847–2855, 2009.

[44] Bagossi P, Horváth G, Vereb G, Szöllösi J, and Tözsér J. *Biophys J.*, 88(2):1354–1363, 2005.

Part II

Results

Chapter 3

Homology and MD Models of the Extracellular Domains

This chapter is based on the published article *"Simulation of homology models for the extracellular domains (ECD) of ErbB3, ErbB4 and the ErbB2–ErbB3 complex in their active conformations"* by JF Franco-Gonzalez, J Ramos, VL Cruz and J Martínez-Salazar, *J. Mol. Model.*, 19(2):931-941, 2013

Abstract

In this work, we provide models based on homology for the following ectodomain (ECD) receptors in their active conformation: ErbB3, ErbB4 and ErbB2/ErbB3 complex. All of them including their own ligand (except for ErbB2), Heregulin-α. We also performed a refinement of the models by molecular dynamics simulations at atomistic scale. We compare the results with a model built for ErbB2 based on crystallographic information and analyze the common features observed among members of the family, namely, the "periscope" movement of the dimerization arm and the "hinge" displacement of subdomain IV.

3.1 Introduction

As mentioned in the Chapter 2, crystal structures for the extended configurations of ErbB1 [1, 2] and ErbB2 [3, 4] are available. Additionally, homology structures for ErbB3 and ErbB4 have been built using as template the crystal structure of 2:2 EGF:EGFR complex (PDB code: 1IVO) [5]. Note that this template lacks the disordered subdomain IV, and homology structures can be obtained without this critical domain. It is only recently that the crystal structure of the 2:2 EGF:EGFR dimer containing subdomain IV has been reported [6]. Therefore, we propose to use this structure as a template to model a more complete structure of the ErbB3 and ErbB4 receptors.

The objective of the present study was to build homology models of the ECDs corresponding to ErbB3 and ErbB4 receptors in their extended conformations along with a model of the back-to-back ErbB2/ErbB3:HRG-α heterodimer.

Water-equilibrated models of the ErbB2, HRGα:ErbB3 and HRGα:ErbB4 complexes were built using as templates the X-ray crystal structure of the ErbB2 / Tzb (PDB code: 1NZ8) and 2:2 EGF:EFGR dimer (PDB code: 3NJP) complexes, respectively, as starting structures, due to the high homology of this receptor family. MD simulations (100 ns) were performed and details of the interactions and intrinsic motions were investigated for each model through Principal Component Analysis (PCA). Additionally, a homology structure of the ECDs of the back-to-back ErbB2/ErbB3:HRGα complex were built and refined through multi-nanosecond MD simulation. To our best of our knowledge, this is the first computational model of this important complex.

3.2 Computational Methods

3.2.1 ECD-Complexes Models

The amino acid sequences of HRG-α, ErbB3 and ErbB4 were obtained from the SwissProt database (P04626,P21860,Q15303). Homology models of the HRG-α:ErbB3 and HRG-α:ErbB4 complexes were built using as a template the X-ray crystal structure of chain A of the 2:2 EFG:EGFR dimer (PDB code: 3NJP) [6]. Unlike the published homology models for HRG-α:ErbB3 and HRG-α:ErbB4 based on the 1IVO structure [5], our template allows the region of sub-domain IV to be included in the homology model. The importance of the orientation of this tight domain has been recognized because of the dependence of intracellular protein kinase domain orientation on sub-domain IV disposition [1, 2].

All models were generated using the PRIME application of the Schrödinger Suite 2011 (Schrödinger, New York, NY). First, pairwise sequence alignments between ErbB3 and ErbB4 and the EFGR template sequences were performed. The BLAST alignment gives 61% and 63% of positives (percentage of residues that

are positive matches to the BLOSUM62 similarity matrix [7]) and 46% and 47% of identities (percentage of residues that are identical between sequences) using 1% and 1% of gaps between the template and ErbB3 and ErbB4, respectively. However, this simple alignment does not assign adequately the secondary structure between the template and the query, as predicted by SSpro program [8]. For this reason, we adopted an alignment algorithm (single template algorithm, STA) that takes into account both secondary structure matching and profile–sequence matching. The Figures 3.8–3.7 in the section 3.6 show the Alignment used for the homology modeling of the Ecto-ErbB2, Ecto-ErbB3, Ecto-ErbB3 and HRG-α proteins, respectively. This procedure minimizes the inaccuracy in a single secondary structure prediction at the expense of increasing the percentage of gaps. This algorithm was designed specially for protein sequences with medium-to-high sequence identity (>25%). The 3D-structures were built by replicating the backbone atom coordinates for the aligned regions and side chains of conserved residues, followed by optimization of the side chains and non-template residues. Finally, the gaps were filled by insertions and final closing of deletions in the alignment. In all cases, the gap length was less than 15. Finally, the homology model coordinates were minimized with Schrödinger [9] using OPLS-2005 with the default threshold of 0.05 kJ/mol as the convergence criteria. The same procedure was used for the homology model of the HRG-α peptide ligand taking as template the EGF peptide of the 3NJP structure. The final HRG-α:ErbB3 and HRG-α:ErbB4 models were built using both sets of coordinates, i.e., receptor and ligand moieties, respectively.

3.2.2 ECD-Erbb2 Model

The initial model of the ErbB2 ectodomain was extracted directly from the 3D-crystal structure deposited in the Protein Data Bank server (PDB code: 1N8Z) [3]. The missing residues N^{124}–A^{132} (sub-domain I), E^{325}–G^{327} (sub-domain II), G^{383}–G^{386} (sub-domain III) and G^{603}–P^{612} (sub-domain IV) were modeled based on homologous sequences using the PRODAT database implemented in Sybyl 8.0 [10]. The loop fragment that affords the best geometric fit as monitored by the homology score and RMS fit, was incorporated automatically into the model [10, 11]. The side chains were built using the rotamer library of Sybyl by taking a scan angle of 30^o and a VDW factor of 0.9. The selected side-chain conformation is the one that presents the fewest bumps on the rest of the molecule. Finally, the structure was relaxed over 2500 steps using the steepest descent minimization algorithm as implemented in GROMACS 4.5.3 [12]

We assume that the pKa of the individual amino acid residues at physiological pH does not change when they are assembled into the protein receptor. Thus, histidine (H) residues were kept neutral while lysine (L) and arginine (R) were protonated and aspartic (D) and glutamic (E) acids were deprotonated. The

resulting total charge for the complex was -10e. The system was then solvated by adding 60717 water molecules and 10 Na^+ ions to yield an electrically neutral system. The system was finally equilibrated in a 2 ns NPT-MD simulation with position restraint for all protein atoms.

3.2.3 Molecular Dynamics Simulations

The OPLS force field [13–15] for protein and SPC model [16] for water were used throughout this work. Short-range repulsion-dispersion interactions were truncated smoothly at 10 Å. The Particle Mesh Ewald (PME) method [17, 18] was used to calculate long-range electrostatic interactions, by means of a maximum grid spacing of 2.5 Å and using fourth-order (cubic) interpolation for the fast Fourier transforms. The temperature was kept constant at 300 K by coupling the protein, the ions and the solvent independently to an external bath using the Berendsen algorithm [19] with a coupling constant of 0.2 ps.

We used isotropic scaling for the pressure (1 bar) and a coupling constant of 1.0 ps and a compressibility of $4.5e10^5$ bar^{-1} when applying the Berendsen algorithm [19]. The dynamics were integrated using the velocity Verlet integrator, with a time step of 2 fs and bonds constrained using the LINCS algorithm [20].

Production dynamics were performed at constant pressure and temperature (NPT ensemble) releasing all constraints on the heavy atoms during 100 ns, and storing the trajectory every 10 ps. All minimizations, restrained and unrestrained MD runs were performed with GROMACS 4.5.3 [12]. Molecular graphics were drawn using the VMD 1.8.7 package [21].

3.2.4 PCA, Hydrogen Bonds and Contact Maps

PCA is a method that takes the trajectory of long MD simulations and calculates the dominant modes in the motion of the molecule. Thus, the conformational space is reduced, resulting in few relevant collective degrees of freedom over which long-range fluctuation can be studied [22, 23]. PCA diagonalizes the covariance matrix of the atom fluctuations from their average trajectory. In this framework, the larger eigenvalues capture the larger fluctuations fraction. The ordering of these eigenvalues gives rise to a small set of modes that capture most of the protein's fluctuation. We performed PCA in order to identify the most relevant motions occurring in the EGFR family. In this work, we make use of the first five eigenvectors, which were projected along the MD trajectory.

Hydrogen bonds (HB) are considered to exist when both distance between the donor (D) and the acceptor (A) is less than 0.30 nm and the DHA (hydrogen-donor-acceptor) angle is less than 30^o.

The contact maps show the smallest distance between any pair of atoms belonging to two different residues. The output is a symmetrical matrix of smallest

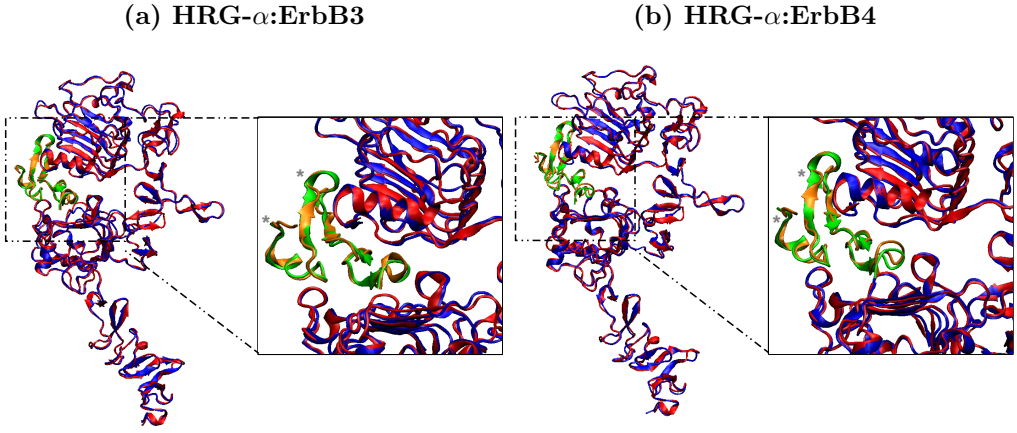


Figure 3.1: Refined Homology Model. **(a)** HRGα(orange):ErbB3(blue) and **(b)** HRGα(orange):ErbB4(blue) superimposed on the X-ray structure EGF(green):EGFR(red) (PDB code: 3NJP). The areas on the right are enlargements showing ligand–receptor interactions.

distances between all residues. Plotting these matrices for different time-frames is a useful tool with which to analyze changes in the structure, HB networks and hydrophobic contacts.

The root-mean square fluctuation (RMSF) for each residue was calculated using the g_rmsf tool from GROMACS. The change in secondary structure elements during the simulation was monitored using the program DSSP (Define Secondary Structure of Proteins) [24].

3.3 Comparative Homology Modeling

3D homology models were built for the ECD domain of HRG-α/ErbB3 and HRG-α/ErbB4 complexes, using as template the X-ray structure corresponding to chain A of the 2:2 EGF:EGFR dimer (PDB code: 3NJP) [6]. The EGFR template has a high sequence similarity of 41% and 42% for the ErbB3 and ErbB4 proteins, respectively. Cysteine residues forming disulfide bonds in both receptor and ligand were well conserved during the alignment process. Figure 3.1 shows the Cα backbone superposition of the refined HRGα:ErbB3 and HRGα:ErbB4 on the X-ray structure of the EGF:EGFR. The root mean squared deviations (RMSD) are 0.8 and 0.7 Å for both ErbB3 and ErbB4 Cα backbones, respectively. As expected, a high degree of structural similarity was observed due to the high sequence homology. Furthermore, all secondary motifs are well captured as seen in Figure 3.1.

The HRG-α ligand has been modeled using the EGF structure in the 3NJP

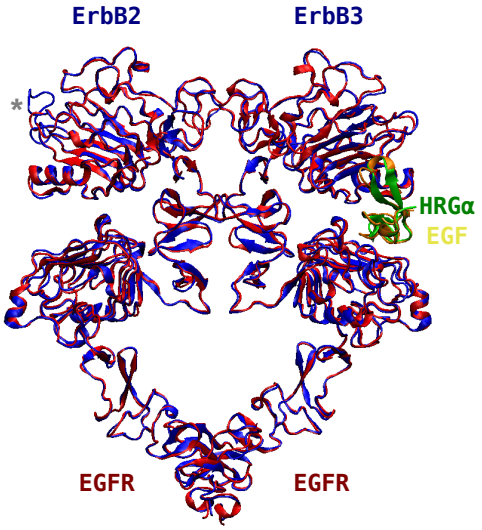


Figure 3.2: ErbB2/ErbB3:HRGα extracellular heterodimer complex (blue for ErbB2 and ErbB3 receptors and green for HRGα ligand) using the $(EGFR:EGF)_2$ homodimer (red and yellow) as template. The asterisk marks the residues from 70 to 114 (flexible loop) where the major differences between the target and the homology model are found

complex as template for both ErbB3 and ErbB4 cases. The sequence identity of the EGF template and HRGα ligand is found to be around 25%. There are a couple of NMR-solved structures for HRGα (PDB code: 1HAF [25] and 1HRF [26]) . However, the EGF ligand in the 3NJP complex is preferred here over the NMR structure in solution, because in principle the ligand–protein interactions are better modeled in the 3NJP structure where the surroundings are taken into account. RMSD values of 3.1 and 3.0 Å were calculated between the EGF Cα backbone atoms in 3NJP and HRGα in ErbB3 and in ErbB4 homology models, respectively. These high values are attributable to different residues in the region of 1-6 (unbounded loop) and 23-28 (flexible loop between β-sheet structures) of EGF and HRG-α ligands (asterisks in the zoomed area of Figure 3.1).

The models were evaluated using Ramachandran plots calculated by the Procheck program. The percentage occupancies for favored, allowed, generously allowed and disallowed regions are collected in Table 3.1. A percentage of 96.1 and 97.3 % of the residues are in favored and allowed regions for the HRGα:ErbB3 and HRGα:ErbB4 models, respectively. After these analyses, the quality of the HRGα:ErbB3 and HRGα:ErbB4 complexes seem to be good enough for further study.

The homology model of the ErbB2:ErbB3:HRGα extracellular complex fitted on the X-ray structure of EGFR:EGF homodimer is shown in Figure 3.2. The RMSD of the Cα backbone atoms are 3.1, 0.8 and 3.0 Å for ErbB2, ErbB3 and HRGα ligand, respectively. The flexible loop made up of the 70-114 residues in ErbB2 is causing the high RMSD value (marked with an asterisk in Figure 3.2). This value is reduced to 0.7 Å once the loop is discarded. Again, the high values of RMSD for the ligand are attributable to the flexible loops discussed above.

Table 3.1: Ramachandran data of HRGα:ErbB3, HRGα:ErbB4 and ErbB2/ErbB3:HRGα homology models calculated with the Procheck software. The 3NJP and 1N8Z X-ray structures are shown as reference. Glycine and proline residues are not taken into account.

	Favored	Allowed	Generously Allow	Disallowed
EGF:EGFR[a]	78.2	19.9	1.1	0.8
ErbB2[b]	82.6	15.3	2.0	0.0
HRGα:ErbB3	72.2	23.9	2.7	1.3
HRGα:ErbB4	72.5	24.8	1.2	1.5
ErbB2/ErbB3:HRGα	72.3	23.5	2.7	1.5

[a] Chain A of the X-ray structure in 3NJP pdb structure
[b] Chain C of the X-ray structure in 1N8Z pdb structure

Following the Ramachandran plot, the 95.8 % of the residues are in favored and allowed regions for the models. Thus, the ErbB2:ErbB3:HRGα extracellular complex is good enough for further MD studies.

3.4 Molecular Dynamics Analysis

We have performed MD simulations of each extracellular system in Table 3.1 (except for the EGF:EGFR complex) for at least 100 ns. The aim of these simulations was to study the stability and the dynamics of the homology models.

3.4.1 ErbB2, HRGα:ErbB3 and HRGα:ErbB4

First, the time evolution of different sub-domains RMSD (Figure 3.3a) was used to track conformational motions along the trajectory. No global conformational switch from the extended to the tethered structure was detected in any case. Thus, the extended conformations remain stable for ErbB2, HRGα:ErbB3 and HRGα:ErbB4 systems as corresponding to active states. Subdomains I and III remain stable along the whole trajectory to values smaller than 2 Å from the crystal structure (ErbB2) or homology models (HRGα:ErbB3 and HRGα:ErbB4). Sub-domains II and IV for ErbB2 are stable with respect to the reference crystal structure, showing only a small deviation of around 3 Å.

On the contrary, subdomains II and IV for the HRG-α:ErbB3 and HRG-α:ErbB4 receptors exhibit deviations in the range of 4-7 and 6-10 Å, respectively. The RMSD profiles for HRGi-α:ErbB3 and HRG-iα:ErbB4 complexes are similar, but the subdomain II deviations, which are larger for HRG-α:ErbB4 (7 Å versus 4-5 Å, blue lines in Figure 3.3a), point to a more movable dimerization arm. In summary, the largest deviations were observed for homology-based structures along the MD trajectories.

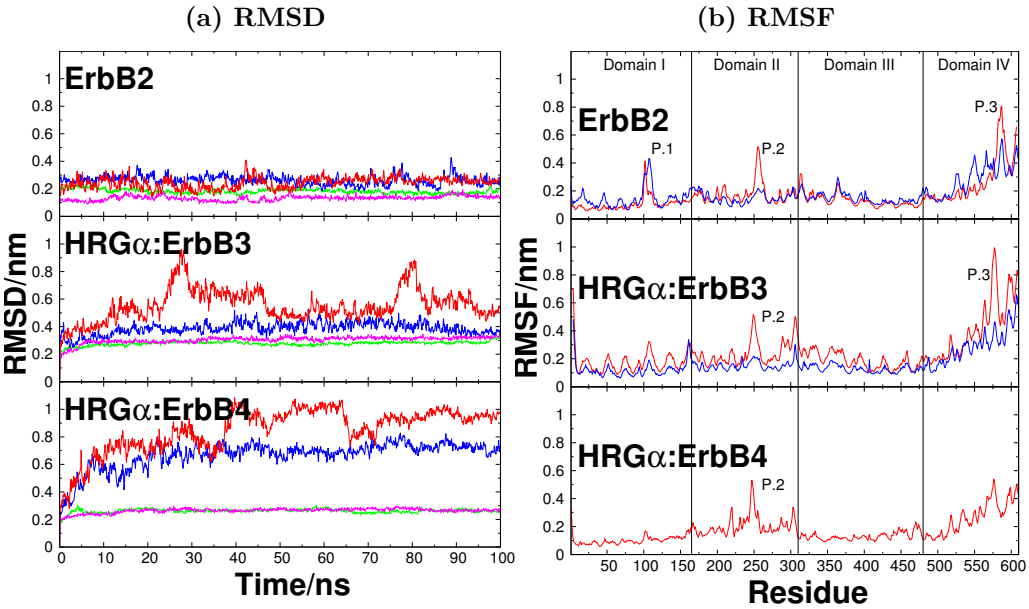


Figure 3.3: **(a)** RMSD of the different domains as a function of time. Subdomains I, II, III and IV are shown as green, blue, magenta and red lines, respectively. **(b)** RMSF as a function of the residue number in backbone atoms. The monomers ErbB2, HRGα:ErbB3 and HRGα:ErbB4 are shown as red lines. The blue lines corresponds to ErbB2 and HRGα:ErbB3 in the heterodimer complex. The last 50 ns of the trajectory were taken in all cases.

Additionally, some remarkable residue moves were observed through RMSF analysis. As shown in Figure 3.3b, the intense peak observed in sub-domain I (peak 1) for ErbB2 corresponds to the motion of a flexible loop (N^{124}–A^{132}). On the contrary, this loop becomes more constrained in ErbB3 and ErbB4 proteins as a consequence of the presence of the HRGα ligand. It should be noted that this loop has not been solved experimentally for the ErbB2 extracellular domain by X-ray techniques, suggesting that the loop is very flexible in that protein [3]. The dimerization arm residues exhibit high RMSF values for all proteins (peak II in Figure 3.3b), showing a periscope-like motion [27]. This motion can assist the formation of dimers between different members of the EGFR family. Finally, domain IV is very mobile with the highest RMSF values (peak 2) in all cases for the proteins alone, corresponding to a hinge motion within this domain. This flexibility could explain the difficulty encountered by other studies in elucidating domain IV by crystallographic techniques on the extended conformations of protein monomers [2, 3]. Although sub-domain II is very flexible in all members of the EGFR family, inspection of both trajectory and RMSD values suggests that ErbB4 is the most flexible.

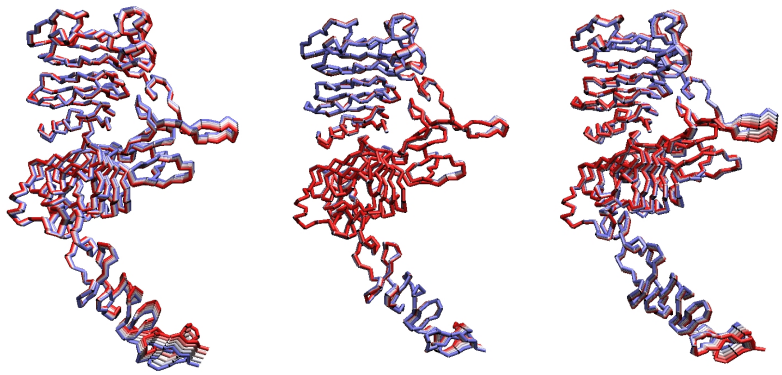


Figure 3.4: Dominant motions represented by the first three eigenvectors in EGFR systems. Several projections are shown in different colors between the two extreme points(blue and red colors). Arrows indicate the direction of motion along the MD simulation for each eigenvector. This figure shows the eigenvectors on the ErbB2 molecule as a example.

Let focus now our analysis on the slowest modes that could provide information on the collective motions of EGFR domains. PCA was performed on the Cα atoms for the three molecules. Despite the fact that PCA was calculated for a single trajectory, some useful qualitative information can be extracted from this analysis. The first three eigenvectors account for Cα 80 % of the motion in ErbB2, HRG-α:ErbB3 and HRG-α:ErbB4 molecules, respectively. The principal motions for the first three eigenvectors of the ErbB2 molecule are shown in Figure 3.4. The motions are similar for both ErbB3 and ErbB4 receptor. They are shown in the Figure 3.11 in the section 3.6. The corresponding videos are also available in the Supplementary Information[1] section of the manuscript by Franco-Gonzalez, et. al [28]. A concerted hinge movement of sub-domain IV is observed clearly for the first PCA eigenvector. The other two eigenvectors show clockwise and counterclockwise torsions of sub-domains II and IV, respectively. Thus, PCA analysis confirms that sub-domains II and IV are very flexible, in agreement with the previous discussion.

Summarizing, MD analysis of the free EGFRs led to the conclusion that great flexibility of domains II and IV exists in the extended structures. It could be hypothesized that this intrinsic flexibility would increase the possibility of forming dimers among different members of the EGFR family.

[1] http://dx.doi.org/10.1007/s00894-012-1613-y

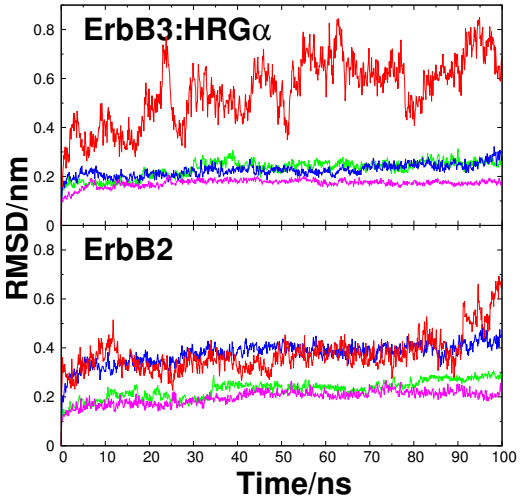


Figure 3.5: RMSD of the domains as a function of time for ErbB2 and HRGα:ErbB3 systems in the heterodimer complex. Domains I, II, III and IV are shown as green, blue, magenta and red lines, respectively.

3.4.2 ErbB2/ErbB3:HRG-α Complex

Global conformation dynamics

The homology model for the ErbB2:ErbB3:HRGα heterodimer is shown in Figure 3.2. We used this model as the initial structure for a 100 ns MD simulation. The overall change in conformation of the heterodimer from the homology structure can be tracked by plotting the RMSD of each sub-domain (Figure 3.5). All sub-domains in the monomers remain stable within less than 0.4 nm from the homology model and they all relax over the first 10 ns. The exception is subdomain II in ErbB3, which has higher RMSD values (ca. 0.6 nm) and large fluctuations, even at the end of the simulation. Moreover, the flexibility of each subdomain of the heterodimer was quantified by calculating the RMSF for Cα atoms in each residue (Figure 3.3b). The fluctuations in subdomains II and IV (peaks 2 and 3) are reduced in the heterodimer (blue lines in Figure 3.3b) as compared to the free monomers (red lines in Figure 3.3b). Therefore, as expected, interactions between the dimerization arms (subdomain II) and domain IV of both ErbB2 and ErbB3 monomers reduce the flexibility of such domains in the heterodimer.

Intermolecular dimerization interfaces

Experimentally, the crystal structure of the $(EGFR:EGF)_2$ homodimer is mediated by intermolecular interactions involving both the dimerization arm in domain II [1, 2, 6] and the C terminus located at subdomain IV [6]. In the present case, the dimerization arm in subdomain II of any monomer (i.e., ErbB2) tends to interact with a region between subdomains I and II of the neighboring receptor. Interactions between residues placed at subdomain IV are also observed

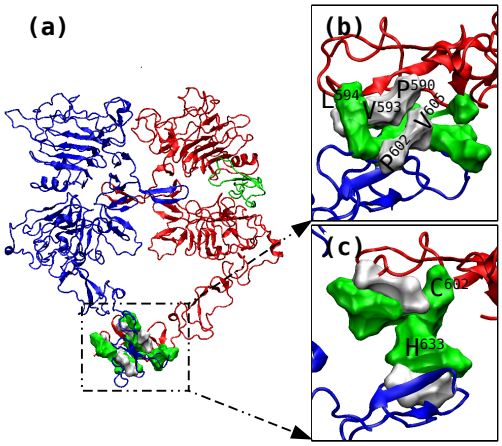


Figure 3.6: (a) Subdomain IV – subdomain IV interface. Interactions in region defined as I (b) and II (c) are zoomed. White surfaces are Hydrophobic contacts and green surfaces are polar contacts. The main residues responsible for each interaction are labeled inside the figure. ErbB2 and ErbB3 receptors are shown as blue and red cartoon models

(see Figure 3.2). The most important non-bonded close contacts between the ErbB2 and ErbB3 are collected in Table 3.2. These contacts take into account salt bridges, HB, aromatic interactions and hydrophobic contacts. As it can be seen in Table 3.2, the subdomain II – subdomain II interface is made up of both hydrophobic and HB interactions in which the 15 residue dimerization arms go through the pocket region of subdomains I and II of the neighboring receptor. However, a persistent interaction with subdomain III of ErbB2 was observed only for the ErbB3 dimerization arm. The subdomain IV – subdomain IV interfaces are shown in Figure 3.6. As can be seen, two regions are kept from the last 50 ns of the trajectory. The first region (Figure 3.6) corresponds to residues A^{599}–L^{609} and S^{587}–E^{596} in ErbB2 and ErbB3, respectively. This region has a marked hydrophobic character with residues P^{602} and V^{605} in the ErbB2 receptor and P^{590}, V^{593} and L^{594} in the ErbB3 receptor.

On the other hand, the second region (Figure 3.6) corresponds to residues C^{631}–C^{635} and C^{617}–G^{623} in ErbB2 and ErbB3, respectively. In this case, the interaction is maintained by the formation of two hydrogen bonds (see Table 3.2).

PCA was carried out on the MD trajectory to identify the most significant cooperative motions of the ErbB2:ErbB3:HRGα complex. The first three eigenvectors account for 75% of the overall motions. These three eigenvectors consist mainly of concerted motions of both subdomains IV (not shown). Remarkably, the motion of subdomain II is more constrained in comparison with the ErbB2 and ErbB3 free receptors.

3.5 Conclusions From This Work

The present work provides a useful collection of homology models for the ECD ErbB receptors updated with the information provided by the latest crystallo-

Table 3.2: Interface non-bonded close contacts (salt bridges, H-bonds,aromatic and hydrophobic interactions) in the ErbB2/ErbB3:HRG-α complex. The domain to which each amino acid belongs is given in parentheses. Only residue-residue contacts with averaged contact area above 20 Å^2 are displayed. The H-bonds reported fulfill the criteria that the donoracceptor distance and angle cutoffs of 3.5 Å and 30^o are maintained during at least 50% of the last 50 ns of the trajectory. The distance in salt bridges is less than 4.0 Å.

	ErbB2	**ErbB3**	**Interaction Type**
Dim. Arm[a]	V^{273}	R^{258}	Hydrophobic
	Y^{275}	C^{301}	H-bond
	D^{278}	T^{108}(I)	Hydrophobic
	F^{280}	R^{267}, Y^{282}	Hydrophobic, aromatic
	E^{281}	A^{304}	Hydrophobic
Dim. Arm[b]	T^{291}, C^{312}(III)	Y^{265}	Hydrophobic, H-bonds
	H^{258}	K^{267}	Hydrophobic
	Q^{58}(I)	L^{268}	H-bonds
	F^{292}, Y^{304}	F^{270}	Hydrophobic, aromatic
	L^{314}(III)	Q^{271}	Hydrophobic
	T^{313}(III)	L^{272}	Not assigned
II – II[c]	G^{224}	N^{224}	H-bond
	P^{233}	Q^{213}	H-bond
	K^{333}	E^{321}	Salt bridge
	R^{352}	E^{273}	Salt bridge
IV – IV[d]	P^{602}, V^{605}	P^{590}, V^{593}, L^{594}	Hydrophobic
	H^{633}	C^{621}	H-bond
	A^{645}	L^{622}	H-bond

[a] Interaction between residues at the dimerization arm of the ErbB2 with the ErbB3 receptor. The dimerization arm in ErbB2 receptor is defined as residues 271–286.
[b] Interaction between residues at the dimerization arm of the ErbB3 with the ErbB2 receptor. The dimerization arm in ErbB3 receptor is defined as residues 261–276.
[c] Persistent hydrogen bonds between residues belonging to subdomain II.
[d] Persistent contacts between residues located at subdomain IV.

graphic structures of suitable templates deposited in the Protein Data Bank. In particular, the model for ErbB3 and ErbB4 receptors includes the structure of sub-domain IV which was absent in previous studies. The quality of the models is proved to be very satisfactory in view of the resulting RMSD differences with the template structures and Ramachandran maps showing suitable backbone torsion angle distributions.

These models were subsequently refined by MD simulation. It is worth mentioning that, according to the RMSD time evolution, the ErbB2 receptor, built from crystallographic data, exhibits the most stable structure along the whole

simulation. A set of common features was found for all the receptors, namely a "periscope" movement of the dimerization arm in domain II, which confirms the findings of previous studies. What is more important is the remarkable flexibility found for sub-domain IV. A "hinge" movement of this domain towards sub-domains II and III was observed in all cases. In this context, PCA reveals that the first eigenvectors are associated to this collective movement.

We have also proposed a model for the interaction of ErbB2 and ErbB3. This complex forms one of the most biologically relevant heterodimers associated with aggressive carcinomas. To the best of our knowledge, the model proposed in our work is the first atomistic scale model for ECD interaction in this heterodimer. The structure presents the expected interaction between the two receptors through the dimerization arms in sub-domain II, which immobilizes these domains with respect to the unbound structures. In addition to this, a weaker interaction through sub-domain IV is also observed. However, the "hinge" movement observed in the separated receptors is also noticeable in the complex in an asymmetric way, being less mobile in ErbB2 than in ErbB3.

The biological consequences of this information are not so evident, and further studies need to be carried out. However, the generation of the ECD ErbBs models presented in this work will serve as an starting point for a systematic study of these important receptors in order to provide clues for the development of more effective therapeutic strategies.

3.6 *Additional Information*

Alignment for the homology models

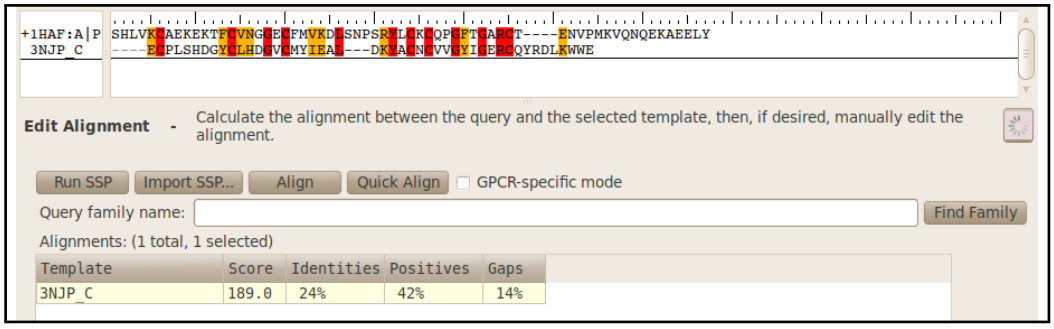


Figure 3.7: Alignment used for the homology modeling of the Heregulin-α.

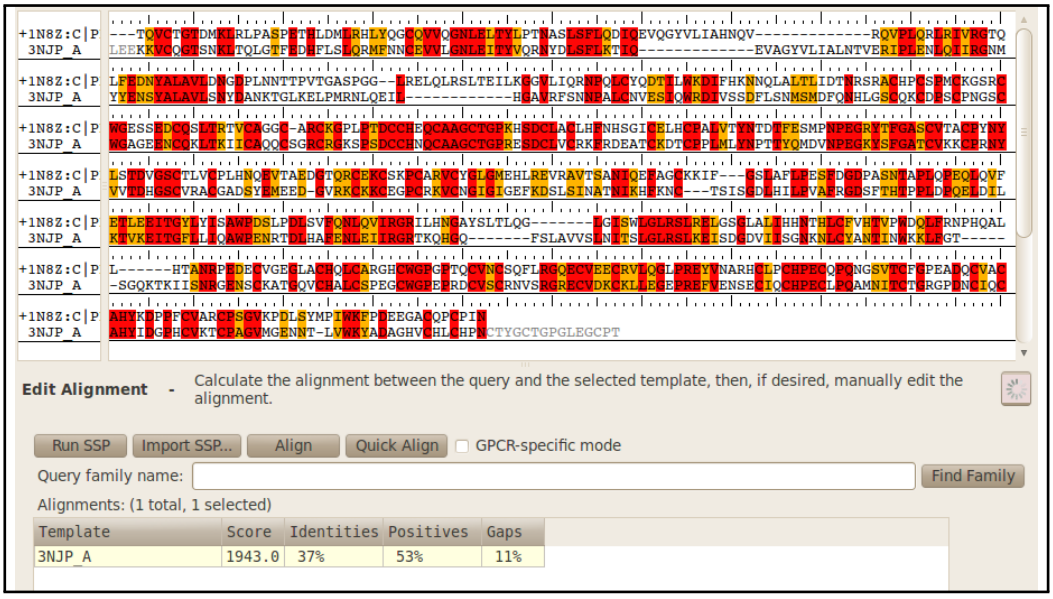


Figure 3.8: Alignment used for the homology modeling of the Ecto-ErbB2 receptor.

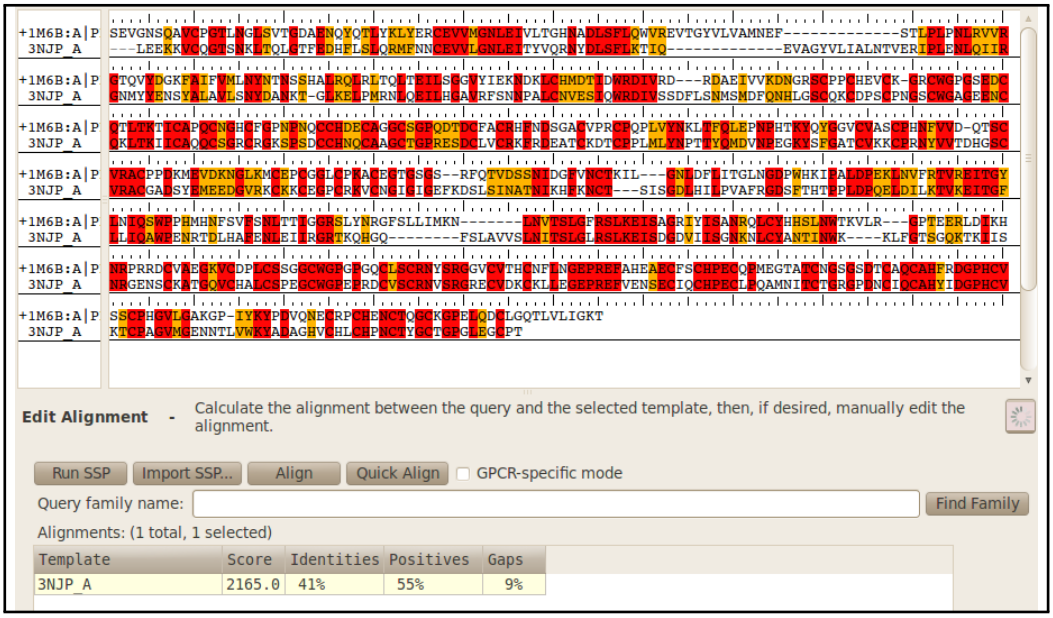


Figure 3.9: Alignment used for the homology modeling of the Ecto-ErbB3 receptor.

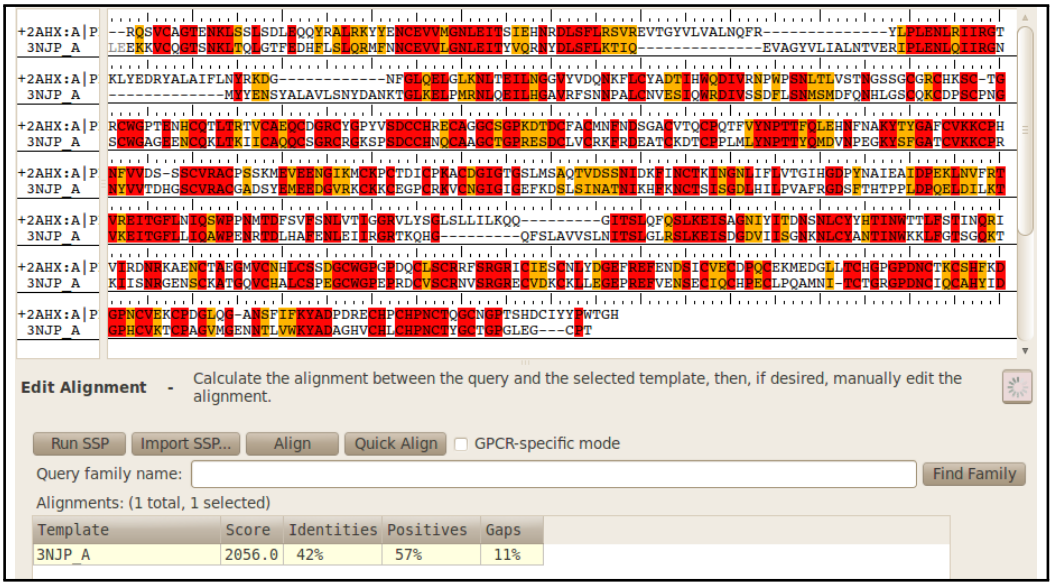


Figure 3.10: Alignment used for the homology modeling of the Ecto-ErbB4 receptor.

Additional PCA eigenvectors

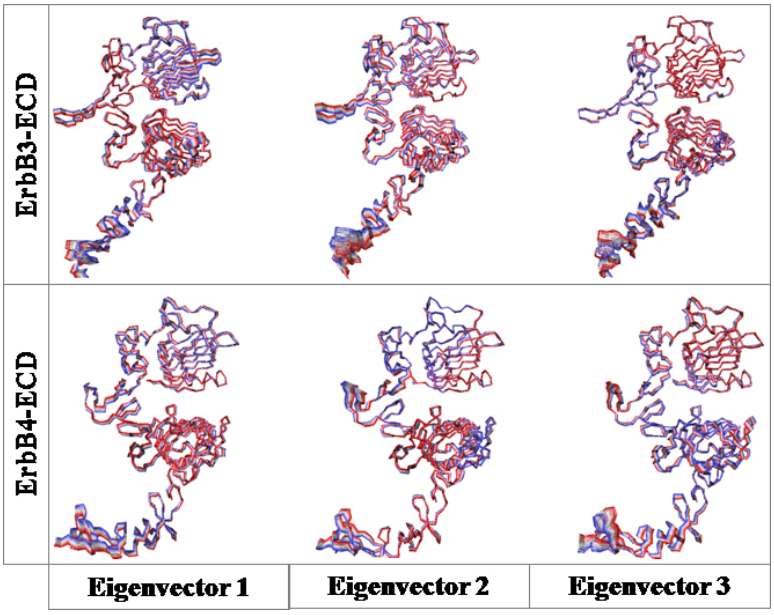


Figure 3.11: First three PCA eigenvectors corresponding to the simulations of the HRGα:ErbB3 and HRGα:ErbB4 ECD receptors

References

[1] Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim J-H, Saito K, Sakamoto A, Inoue M, Shirouzu M, and Yokoyama S. *Cell*, 110(6):775–787, 2002.

[2] Garrett TPJ, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Zhu H-J, Walker F, Frenkel MJ, Hoyne PA, Jorissen RN, Nice EC, Burgess AW, and Ward CW. *Cell*, 110(6):763–773, 2002.

[3] Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, and Leahy DJ. *Nature*, 421(6924):756–760, 2003.

[4] Franklin MC, Carey KD, Vajdos FF, Leahy DJ, de Vos AM, and Sliwkowski MX. *Cancer Cell*, 5:317–328, 2004.

[5] Luo C, Xu L, Zheng S, Luo X, Shen J, Jiang H, Liu X, and Zhou M. *Proteins*, 59(4):742–756, 2005.

[6] Lu C, Mi n Z, Grey MJ, Zhu J, Graef E, Yokoyama S, and Springer TA. *Mol. Cell Biol.*, 30(22):5432–5443, 2010.

[7] Henikoff S and Henikoff JG. *Proc. Natl. Acad. Sci. USA*, 89(22):10915–10919, 1992.

[8] Cheng J, Randall AZ, Sweredoski MJ, and Baldi P. *Nucleic Acids Res*, 33(suppl 2):W72–W76, 2005.

[9] Schrödinger. LLC. Maestro, version 9.2, 2011. New York, NY.

[10] Tripos International. Sybyl 8.0. St. Louis, MO.

[11] Dayhoff M, Schwartz R, and Orcutt B. *Atlas of protein sequence and structure*. National biomedical research foundation, Washington DC, 1978.

[12] Hess B, Kutzner C, van der Spoel D, and Lindahl E. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[13] Jorgensen WL and Tirado-Rives J. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.

[14] Jorgensen WL and Tirado-Rives J. *Proc. Natl. Acad. Sci. USA*, 102(19):6665–6670, 2005.

[15] Lawrence CP and Skinner JL. *Chem. Phys. Lett.*, 372(5-6):842–847, 2003.

[16] Berendsen H, Postma J, van Gunsteren W, and Hermans J. *Intermolecular forces*, chapter Interaction models for water in realtion of protein hydration, pages 331–342. Reidel, Dordrecht, 1981.

[17] Darden T, York D, and Pedersen L. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[18] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[19] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[20] Hess B, Bekker H, Berendsen HJC, and Fraaije JGEM. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[21] Humphrey W, Dalke A, and Schulten K. *J. Mol. Graph.*, 14(1):33–38, 1996.

[22] Amadei A, Linssen ABM, and Berendsen HJC. *Proteins Struct. Funct. Bioinforma.*, 17(4):412–425, 1993.

[23] Amadei A, Linssen AB, de Groot BL, van Aalten DM, and Berendsen HJ. *J. Biomol. Struct. Dyn.*, 13(4):615–625, 1996.

[24] Kabsch W and Sander C. *Biopolymers*, 22(12):2577–2637, 1983.

[25] Jacobsen NE, Abadi N, Sliwkowski MX, Reilly D, Skelton NJ, and Fairbrother WJ. *Biochemistry*, 35(11):3402–3417, 1996.

[26] Nagata K, Kohda D, Hatanaka H, Ichikawa S, Matsuda S, Yamamoto T, Suzuki A, and Inagaki F. *EMBO J.*, 13(15):3517–3523, 1994.

[27] Fuentes G, Scaltriti M, Baselga J, and Verma C. *Breast Cancer Res.*, 13(3):R54, 2011.

[28] Franco-Gonzalez JF, Ramos J, Cruz VL, and Martinez-Salazar J. *J. Mol. Model.*, 19(2):931–941, 2013.

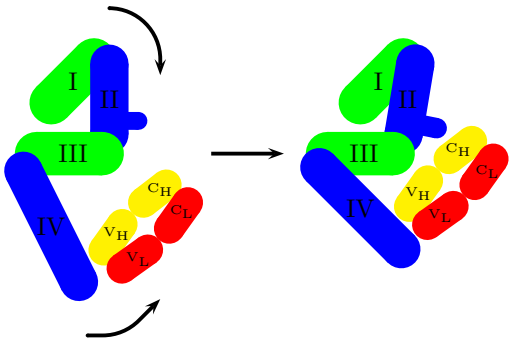
Chapter 4

ECD-ErbB2/Trastuzumab-Fab Complex

This chapter is based on the published article *"Conformational Flexibility of the ErbB2 Ectodomain and Trastuzumab Antibody Complex as Revealed by Molecular Dynamics and Principal Component Analysis"* by JF Franco-Gonzalez, J Ramos, VL Cruz and J Martínez-Salazar, *J. Mol. Model.*, 19(3): 1227-1236, 2013

Abstract

As a successful therapeutic treatment against breast cancer the monoclonal antibody Trastuzumab is used. Its interaction with the ECD-ErbB2 is well-known. A better understanding of the detailed structure of the receptor-antibody interaction is indeed of prime interest for the design of more effective anticancer therapies. In order to discuss the flexibility of the complex ECD-ErbB2/Trastuzumab, a multi-nanosecond molecular dynamics simulations together with an analysis of fluctuations, through a PCA of this system are shown in this chapter.



4.1 Introduction

Several computational studies based on computer simulations have tackled the structure and interactions between transmembrane ErbB2 domains in lipidic bilayer models [1–6] and in tyrosine kinase domain activation [7–9]. However, computational studies of the interaction of ErbB2 ectodomain (ECD-ErbB2) with antibodies, such as Trastuzumab (Tzb), are less considered. Wang et al. have studied the binding regions of ECD-ErbB2 with inhibitory (Tzb) and non-inhibitory (HF) monoclonal antibodies using a combination of site-directed mutagenesis, docking and short molecular dynamics simulations. They concluded that the inhibitory Tzb antibody binds to subdomain IV (C-terminal region) of the ECD and that the non-inhibitory HF antibody recognizes subdomain II (N-terminal region) [10]. In other study, the 3D structure of an autoinhibitor (herstatin)/ECD-ErbB2 complex has been proposed using molecular docking methods. The binding site of herstatin of the ECD-ErbB2 domain was proposed to be at the S1 subdomain (here subdomain II). That observation was verified by inmunoprecipitation, confocal microscopy and fluorescence resonance energy transfer experiments [11]. Very recently, Fuentes et al. have published a 20 ns MD study and a fluctuation analysis of the interaction between ErbB2 and a combination of Tzb and pertuzumab antibodies [12]. Their simulations shed light on two important aspects of the interaction: on one hand, the fluctuations in subdomain II are enhanced by the Tzb binding, and on the other hand, the existence of a cooperative mechanism between these two antibodies and the ErbB2 ECD that could avoid the homo and heterodimerization of ErbB2 with other members of the EGFR family.

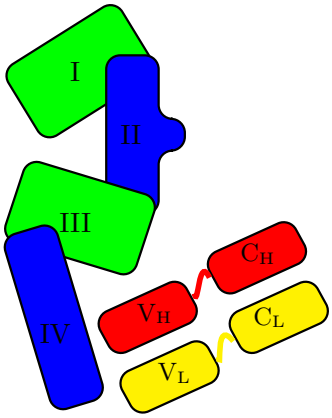


Figure 4.1: Schematic representation of the different domains in the ECD-ErbB2/Trastuzumab-Fab complex. ECD-ErbB2 receptor is shown in green (I and III subdomains) and blue (II and IV subdomains), whereas Tzb, the variable (V) and the conserved domain (C) are represented in yellow (light chain) and red (heavy chain)

In our work we performed a long 170 ns MD simulations of the ECD-ErbB2 /Tzb complex to elucidate details of the interaction between its components using as starting point the x-ray crystal structure [13]. Additionally, an analysis of the large scale fluctuations has been performed using the PCA analysis. It should be mentioned here the experimental finding that the variable Tzb domains bind to subdomain IV (juxtamembrane domain) of ECD-ErbB2. These interactions have been largely conserved along the MD simulation. However, large fluctuations are

observed which allow the formation of novel contacts between the dimerization arm of subdomain II ErbB2 and the Tzb residues in the constant domain. To our best knowledge this interaction has not been yet reported.

4.2 Computational Methods

Structure Modeling

The initial model of the ECD-ErbB2 and Tzb-fab (ECD-ErbB2/Tzb) was directly taken from the 3D crystal structure deposited in the Protein Data Bank server (PDB code: 1N8Z) [13]. A schematic representation can be seen in Figure 4.1. The missing residues 102-110, 303-305, 361-364 and 581-590 have been modeled based on homologue sequences using the PRODAT database implemented in Sybyl 8.0 [14]. The loop fragment that gave the best geometric fit, based on the homology score and RMS fit, was automatically incorporated into the model [14, 15]. The side chains were built residue by residue, one at a time, using the rotamer library of Sybyl using a scan angle of 30^o and VDW factor of 0.9. The selected side chain conformation for each modeled residue is the one that presents the fewest VDW contacts with the rest of the molecule. Finally, the structure was relaxed for 2500 steps using the steepest descent minimization algorithm as implemented in GROMACS 4.5.3 [16]. The Procheck analysis of the added fragments reports a Ramachandran plot with the following statistics: 64.3 % in most favored regions, 21.4 % in additional allowed regions, 14.3 % in generously allowed regions and 0 % in disallowed regions. From that point of view the structure of the added fragments seems to be quite reasonable.

We assumed that the pKa of the individual amino acid residues at physiological pH does not change when assembled into the protein receptor. Thus, histidine (H) residues remained neutral; lysine (L) and arginine (R) were protonated and aspartic (D) and glutamic (E) acids were deprotonated. The resulting total charge for the complex was -10 e units. The system was solvated by 60717 water molecules and 10 Na^+ ions have been added to yield an electrically neutral system. Periodic boundary conditions were applied along the three dimensions. The initial rectangular box lengths were 13.9 nm, 12.5 nm and 11.5 nm respectively. The system was equilibrated in a 2 ns NPT-MD simulation with position restraint for all protein atoms.

Molecular Dynamics

The OPLS force field [17–19] for protein and the SPC model [20] for water were used along the whole work. Short range repulsion-dispersion interactions were smoothly truncated at 10 Å. The Particle Mesh Ewald (PME) method [21, 22] was used to calculate long range electrostatic interactions, with a maximum grid

spacing of 2.5 Å and using fourth-order (cubic) interpolation for the fast Fourier transforms. The temperature was kept constant at 300 K by coupling the protein, the ions and the solvent independently to an external bath using the Berendsen algorithm [23] with a coupling constant of 0.2 ps.

We used isotropic scaling for the pressure (1 bar). A coupling constant of 1.0 ps and a compressibility of $4.5e10^{-5}$ bar^{-1} were used in the Berendsen algorithm [23]. The dynamics were run using the velocity Verlet integrator, with a time step of 2 fs and bonds constrained conditions using the LINCS algorithm [24].

Production dynamics was performed at constant pressure and temperature (NPT ensemble) releasing all constraints on the heavy atoms during 170 ns and accumulating the trajectory frames every 10 ps. All minimizations, restrained and unrestrained MD runs were performed with GROMACS 4.5.3 [16]. Molecular graphics have been drawn using the VMD 1.8.7 package [25].

PCA analysis, Hydrogen Bonds and Contact Maps

With PCA analysis the configurational space is reduced, containing few relevant collective degrees of freedom in which long range fluctuation can be studied [26, 27]. A PCA diagonalizes the covariance matrix of the atom fluctuations from their average trajectory. In this framework, the larger eigenvalues correspond to eigenvectors which explain most of the variance of the atomic fluctuations. The ordering of these eigenvalues gives rise to a small set of modes that capture most of the proteins fluctuations. We have performed a PCA analysis in order to identify the lowest frequency motions occurring in the ECD-ErbB2/Tzb complex. Along this work, we make use of the first three eigenvectors, which were projected along the MD trajectory. The g_covar and g_anaeig tools in the GROMACS package were used to perform the PCA analysis.

Hydrogen bond (HB) is considered to exist when both distance between the donor (D) and the acceptor (A) is less than 0.30 nm and the hydrogen-donor-acceptor (HDA) angle is lower than 30^o. The contact maps show the smallest distance between any pair of atoms belonging to two different residues. The output is a symmetrical matrix of smallest distances between all residues. Plotting these matrices for different time-frames is useful to analyze changes in the structure, and particularly hydrogen bond networks and hydrophobic contacts. The root-mean square fluctuation (RMSF) for each residue has been calculated using the g_rmsf tool from GROMACS. The change of secondary structure elements during the simulation was monitored using the program define secondary structure of proteins (DSSP) [28].

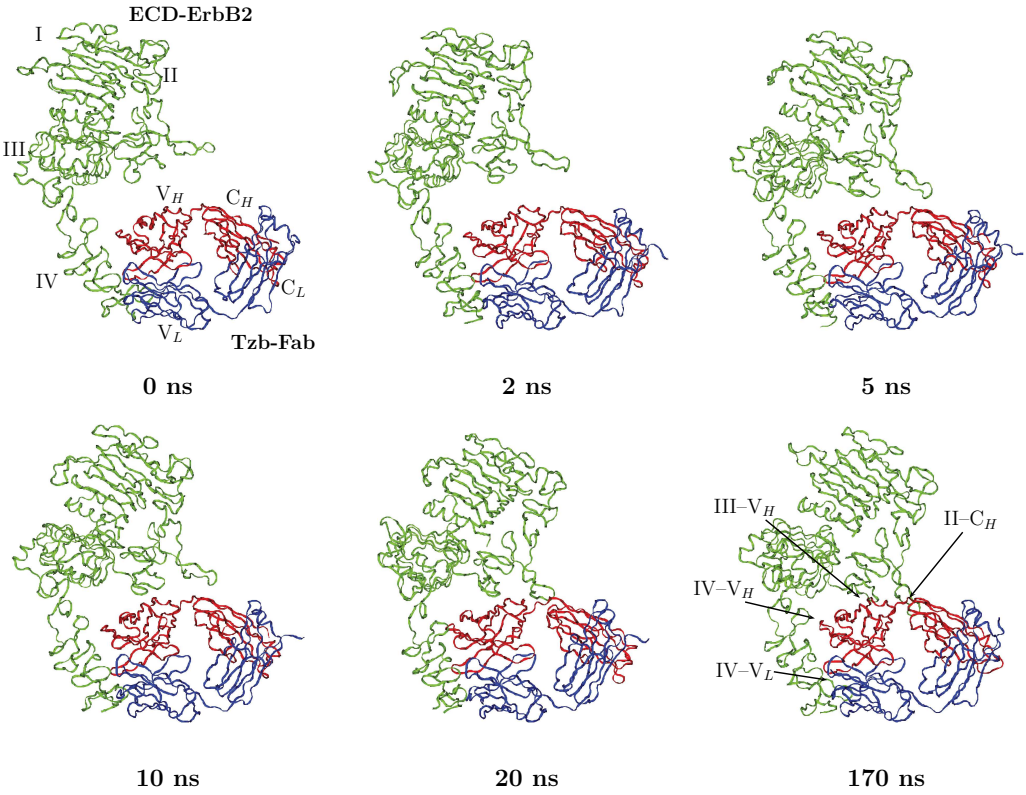


Figure 4.2: Selected structures of the ECD-ErbB2/Tzb complex along the MD trajectory. The ECD-ErbB2 protein and Tzb-fab are shown as yellow and blue/red ribbons, respectively. The configuration at t=0ns corresponds to the x-ray elucidated structure

4.3 Stability Analisys of Full Trajectory

Figure 4.2 shows the time evolution of the interaction in the ECD-ErbB2/Tzb-fab complex from the MD simulations. At t=0 ns, the initial conformation matches to the x-crystal solved structure. It is relevant to mention that the secondary structures are well conserved along the whole MD trajectory (see DSSP analysis section in the section 4.8). Furthermore, the interaction values between the ECD-ErbB2 subdomain IV and V_H/V_L domains of Tzb, which have been experimentally reported, are well preserved during the whole MD trajectory (as will be discussed below). However, after 20 ns, the ECD-ErbB2 subdomain II (dimerization arm) makes contact with the C_H subdomain of the Tzb. Later on, after 100 ns (not shown in Figure 4.2), less important contacts between III and V_H subdomains are found. Most likely, these interactions are a consequence of the driven-contacts between II and the C_H domain.

The time evolution of the Cα-RMSD (root mean square deviation) for the

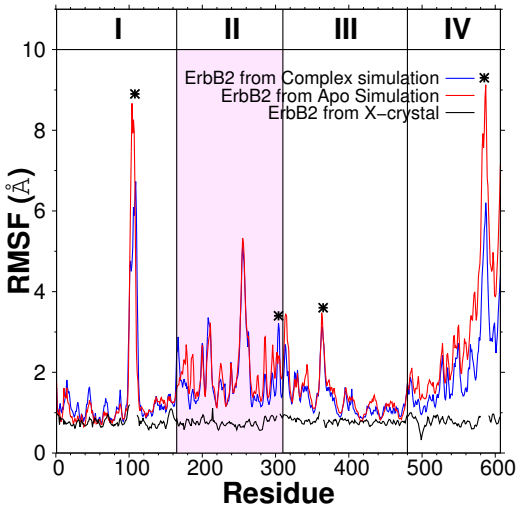


Figure 4.3: ErbB2 residue fluctuations along the MD. Root mean square fluctuation (RMSF) of backbone atoms of ECD-ErbB2 residues from the initial structure (blue line), the apo-protein case (red line) and from the B factors (black line) using the formula RMSF=$(3B/8\pi^2)^{1/2}$. Asterisk stands for missing loops in the x-ray structure 1N8Z. The shaded area indicates residues in subdomain II

ECD-ErbB2 receptor has been used to track the equilibration and any possible reorganization of present domains in the whole complex (see Stability analysis section in the section 4.8). The Cα-RMSD relaxes over the first 70 ns to a value of around 0.41±0.03 nm (averaged over the interval 70-170 ns). These structural changes are mainly due to subdomains IV and II of the ECD-ErbB2, showing the largest values (Cα-RMSD$_{(70ns-170ns)}$ 00.27±0.02 and 0.31±0.04 for subdomains IV and II, respectively). On the other hand, Cα-RMSD values for the subdomains I and III keep stable around 0.12±0.02 nm along the whole trajectory. As shown in Scheme 4.1 and Figure 4.2, the subdomains II and IV are more exposed to the interaction with the V_L and C_H domains of the antibody structure, respectively.

The ECD-ErbB2 residue RMSF values for the x-ray structure (calculated from B factors of the 1N8Z file as $(3B/8\pi^2)^{1/2}$) and for the MD simulations are presented in Figure 4.3. As can be seen in the MD simulations, the largest fluctuations are concentrated in the missing loops of the x-ray structure (marked with an asterisk). These large fluctuations are compatible with the fact that these residues cannot be solved in the crystalline structure [13]. Furthermore, several peaks corresponding to the dimerization arm residues (subdomain II) show large fluctuations (shaded area in Figure 4.3), which are well-suited with the movement discussed above. This is in agreement with similar fluctuations of subdomain II recently observed in shorter MD simulations (20 ns) of the ECD-ErbB2/Tzb complex using a different force field [12]. However, these residue fluctuations in the subdomain II are not observed in the experimental data. This may indicate that the subdomain II region in the x-ray structure is rather constrained due to the crystalline packing [29, 30]. We will come back to this difference in a subsequent section. We have included in Figure 4.3 the RMSF values of the ECD-ErbB2 simulation in water to compare with the values corresponding to the

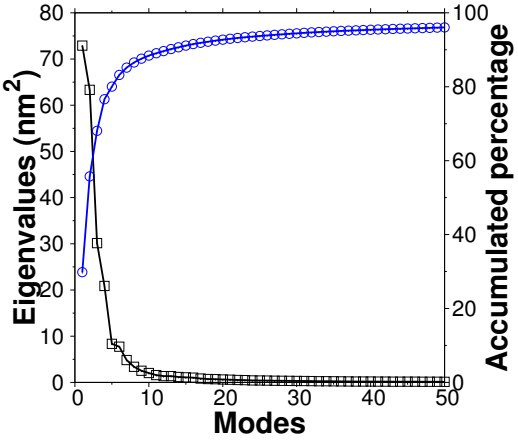


Figure 4.4: PCA analysis. Eigenvalues (squares, scale on left side) and accumulated percentage (circles, scale on right side) of the first 50 PCA modes

ErbB2/Tzb complex. As can be observed, the fluctuations in the dimerization arm are very similar in both cases. The main difference between both RMSF profiles corresponds to subdomain IV, which shows a higher flexibility in the apo-protein case.

4.4 PCA Analysis

The PCA allows the projection of the complex protein dynamics on a set of collective modes which can be ordered from the largest to smallest contributions of the protein fluctuation variance, as measured by the eigenvalues of the covariance matrix [26, 27]. The largest eigenvalue corresponds to the slowest motion, and so forth.

The contributions to the motion for the 50 first collective modes are shown in Figure 4.4. The major contribution to the collective motion is given by the first nine modes with 90 % of the total protein fluctuations. Modes 1 and 2 contribute with 30 and 26 % of the overall motion with fluctuations of 73 and 63 nm^2, respectively. Mode 3 gives 12 % with a fluctuation value of 30 nm^2. These three first modes account for ∼70 % of the total protein fluctuations and only the nine first eigenvalues have a value greater than 1 nm^2 (Kaiser criterion discarding eigenvalues below 1) [31] accounting for a 90 % of the total fluctuations.

The three first modes along with the RMSF residue values for each of them are shown in Figure 4.5. In the first mode, part of the subdomain II in ErbB2 translates in a concerted way to domain C_H in the Tzb protein. Upon this movement, large fluctuations of the flexible loops (102-110 and 581-590, marked with asterisk) in the subdomain I and IV are also observed (Figure 4.5b, black line). Mode 2 is dominated by the movements of the loop 102-110 in subdomain I with some minor contributions from a concerted rotational approach of the II and C_H domains of the ErbB2 and Trastuzumab proteins (Figure 4.5b, red line).

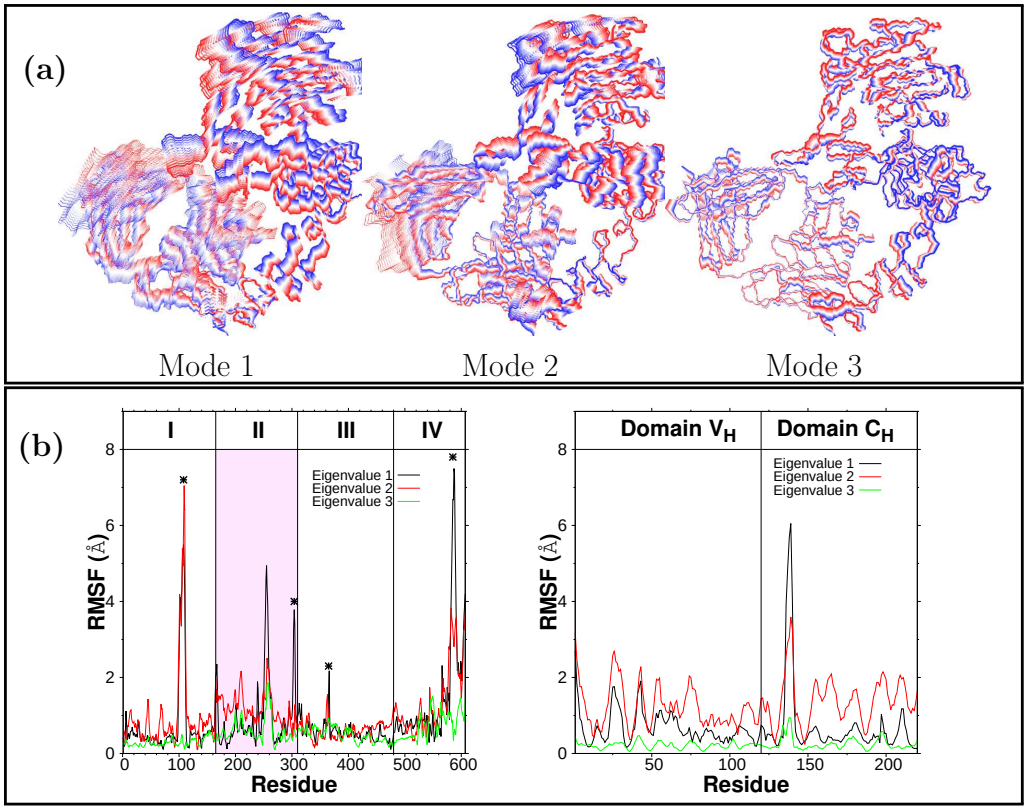


Figure 4.5: Eigenvectors from the PCA analysis. **(a)** The motion of the three first principal components is overlaid sequentially. Red and blue colors represent large and low-amplitude mode, respectively. The dimerization arm and the Transtuzumab domains exhibit the largest localized motions assisting to the interaction between both regions (see text for details). **(b)** RMSF of residues for each mode, 1 (black line), 2 (red line) and 3 (yellow line) in the ERBB2 and Tzb proteins. Asterisk stands for missing loops in the x-ray structure 1N8Z. The shaded area indicates residues in subdomain II of the ErbB2.

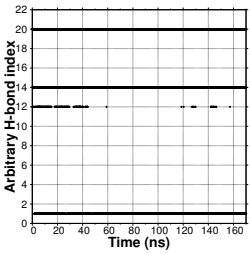
Finally, in mode 3, only a torsional combined motion of the II and C_H domains contributes significantly to the overall fluctuation.

These eigenvalues involve large motions of the subdomain II and C_H motions which can be confirmed by visualizing the distances between the center of mass of the different domains in ErbB2 and Trastuzumab moieties. The approach of II and C_H domains is evidenced by a decrease of more than 2 nm of the center of mass distance (CMD). On the other side, the CMD between subdomain IV in ErbB2 and Tzb domains are kept nearly constant throughout the full dynamics and close to those found in the x-ray structure.

In summary, these principal component eigenvectors show a hinge motion in which subdomain II and C_H domains approach each other, allowing the formation

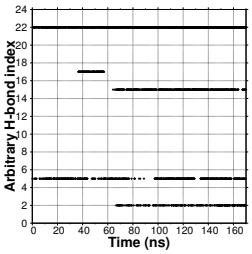
(a)

Index	Donor	Acceptor	Occupancy (%)
1	ARG577(IV)	ASP102(V_H)	99.8
12	LYS593(IV)	GLY103(V_H)	3.1
14	ARG50(V_H)	GLU580(IV)	100
20	ARG59(V_H)	ASP560(IV)	95.4



(b)

Index	Donor	Acceptor	Occupancy (%)
2	MET585(IV)	TYR49(V_L)	28.2
5	GLN602(IV)	ASN30(V_L)	50.4
15	TYR49(V_L)	LEU586(IV)	53.1
22	THR94(V_L)	ASP560(IV)	87.3



(c)

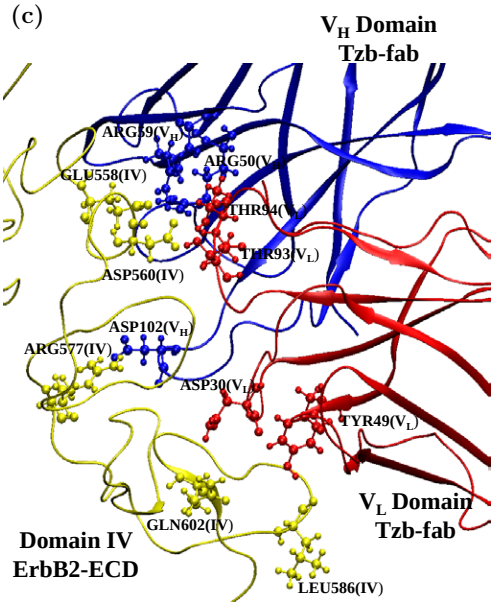


Figure 4.6: Hydrogen bond network between antibody and domain IV. **(a)** Hydrogen bonds between IV and V_H domains, **(b)** Hydrogen bonds between IV and V_L domains and **(c)** Snapshot at 170 ns of the hydrogen bond networks between IV (yellow), V_H (blue) and V_L (red) domains

of some interactions between the dimerizarion arm in the subdomain II of ErbB2 protein and the Tzb residues located in the C_H domain. These interactions will be discussed in the next section where a detailed analysis of the inter-domain hydrogen bonds and electrostatic interactions between ErbB2 and Tzb proteins is given. Movies containing the motion of the first three eigenvectors can be seen in the Supplementary Information[1] section of the manuscript by Franco-Gonzalez, et. al [32].

4.5 Subdomain IV – Tzb Interactions.

Hydrogen bonds have been characterized according to their residence times during the MD trajectory using the distance and angle criteria defined in the Compu-

[1] http://dx.doi.org/10.1007/s00894-012-1661-3

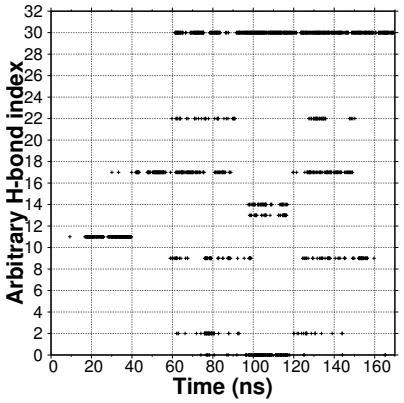
tational methods section. Firstly, we study the described interactions between subdomain IV and Tzb and compare them with the experimentally available data. Figure 4.6a shows the time evolution of the hydrogen bonds which describes the interaction of the subdomain IV of the ECD-ErbB2 protein and the domain V_H of the Tzb-fab antibody. According to the calculations, three pairs are maintained from the starting structure during the whole simulation, Arg577(IV)–Asp102(V_H), Arg50(V_H)–Glu558(IV) and Arg59(V_H)-Asp560(IV) with occupancies close to 100 %. These interaction pairs can be better considered as salt bridges between charged acidic and basic residues. Similarly, the hydrogen bonds between the subdomain IV of the ECD-ErbB2 protein and the domain V_L of the Tzb-fab antibody are shown in Figure 4.6b. As can be seen, hydrogen bonds are more labile (small residence time) and they are mainly formed by two pairs, Asn30(V_L)–Gln602(IV), and Thr94(V_L)-Asp560(IV). Furthermore two new H-bonds are alternatively formed by the pair Tyr49(V_L)–Leu586(IV) and Met589(IV)–Tyr49(V_L) after 80 ns. All these interactions (Figure 4.6c) are in close agreement with the experimental data reported by Cho et al. [13], who showed electrostatic interactions between the 557–560 and 593–603 loops of the ErbB2 ectodomain and the Tzb antibody. Hydrophobic interactions have also been described between the loop 570–573 located in the subdomain IV of the ErbB2 receptor and the Tzb-fab antibody in the same experimental work [13]. The hydrophobic interactions were analyzed with the help of the contact map analysis utility which is available at the SPACE server for protein structure analysis (http://ligin.weizmann.ac.il/cma/) [33]. The hydrophobic contacts were selected from the list provided by CMA (Contact Map Analysis) with two conditions, namely, the contact surface should be above 0.4 nm^2 and the two involved residues should be hydrophobic. Along the MD simulation, the main hydrophobic interactions between these domains are effectively located between the loop 570–573 of the ErbB2 receptor and the hydrophobic CDR3 loop of the V_H domain (loop 101–110) in the Tzb-fab. In addition, hydrophobic interactions between the 570–573 loop and the CDR3 loop of the V_L domain (loop 93–99) are also observed. Thus, it can be concluded that these hydrogen bond interactions and hydrophobic contacts are very stable, maintaining their distances along the full simulated trajectory.

4.6 Subdomain II – Tzb Interactions.

Based on the global structural changes observed in the previous section for the subdomain II, we have investigated possible hydrogen bond interactions and electrostatic contacts with the antibody domains. Figure 4.7 shows the hydrogen bond network between subdomain II and C_H region. It can be observed that no hydrogen bonds are presented at the beginning of the simulation. Later on, some

(a)

Index	Donor	Acceptor	Occupancy (%)
0	ASN253(II)	ASN211(C_H)	15.5
2	THR256(II)	ASN211(C_H)	13.5
9	LYS208(C_H)	GLU258(II)	9.0
13	ASN211(C_H)	THR251(II)	4.4
14	ASN211(C_H)	THR252(II)	6.7
17	ASN211(C_H)	THR256(II)	13.5
22	ASN211(C_H)	SER259(II)	5.0
30	LYS213(C_H)	ASP255(II)	73.2



(b)

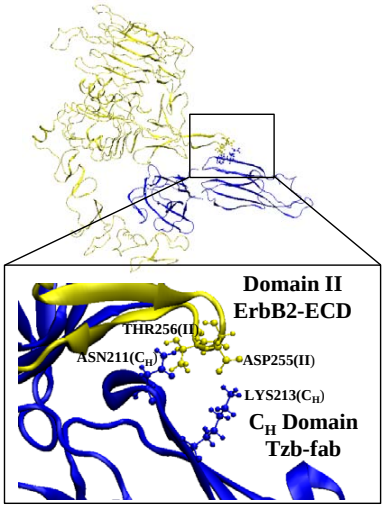


Figure 4.7: Hydrogen bond network between antibody and domain II. **(a)** Hydrogen bonds between II and C_H domains, **(b)** Snapshot at 170 ns and zoom of the hydrogen bond networks between II (yellow) and C_H (blue) domains

residues start forming hydrogen bonds, being the pairs Asn211(C_H)–Thr256(II) and Lys213(C_H)–Asp255(II) the most stable along the trajectory. The occupancy (fraction of time in which the H-bond is formed) of these H-bonds are 27 % and 74 %, respectively, for the last 70 ns. Predominantly, a salt bridge between Asp73(V_H) and Lys346(III) is formed at distances fluctuating between 3.0 and 4.0 Å. The formation of these H-bonds and electrostatic contacts might contribute to stabilize the structure resulting from the hinge motion described above in the PCA analysis, where we stated that the subdomain II has a large RMSD drift and RMSF values closer to the CH domain of the antibody.

As shown in the previous paragraph, the dimerization arm is at least as flexible as loops 102–110 and 581–590, which are missed in the x-ray structure indicating its high flexibility (Figure 4.3). However, the dimerization arm structure has been experimentally solved. A close inspection of the crystallographc packing can help to interpret this issue. Figure 4.8 illustrates two views of the monomer surroundings in the unit cell of the x-ray structure. It can be seen that the dimerization arm is located close to the back-position of a different ErbB2/Tzb monomer and interacting with the Trastuzumab domain of another different monomer. Thus,

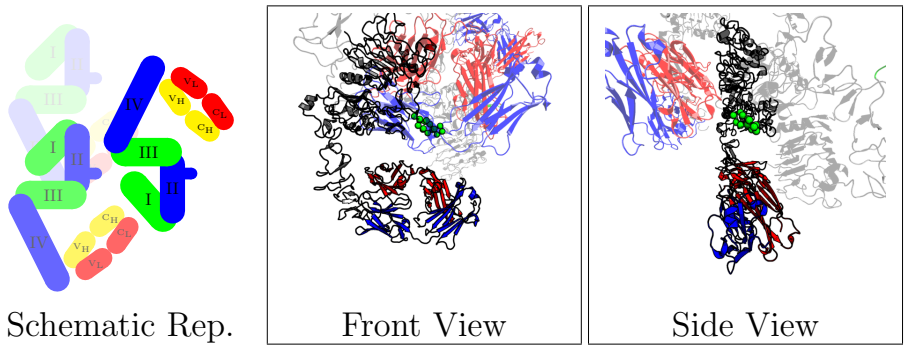


Figure 4.8: Crystal packing of the ErbB2/Tzb complex. On the left Schematic representation of the crystal packing of three ErbB2 complexes in 1N8Z. At center and right Front and Side views of the 1N8Z crystal packing. Cartoons shows an ErbB2/Trastuzumab complex along with the nearest spatial neighbor domains represented with lines. The dimerization arm is shown as red VdW spheres

both the dimerization arm and the hinge motions are hindered by the closed presence of other monomers in the crystal structure. In our simulations, the lack of other monomers shows evidence of the intrinsic flexibility of the complex ErbB2/Tzb, being likely stabilized by H-bond and electrostatic interactions between both components.

On the other hand, the binding free energy has been calculated according to the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method. The "in-silico" binding energy ΔG_{bind} is quite large (-285.0 kcal mol^{-1}) in comparison with the reported experimental values between -12.4 and -14.0 kcal mol^{-1} [34, 35]. In this sense, Fuentes et. al. [12] reported an "in-silico" value of ΔG_{bind} -1144.6 kcal mol^{-1} using molecular mechanics-generalized-Born surface area (MM-GBSA) approximation without entropic terms. Thus, it seems clear that entropic terms are needed in order to improve the binding energy between the apo-ErbB2 protein and the Tzb ligand. Details of the MM-PBSA calculation are given in the section 4.8 in the section 4.8.

4.7 Conclusions From This Work

A molecular dynamics and PCA study of the flexibility of the complex between the extracellular domain (ectodomain) of the ErbB2 receptor and the Trastuzumab antibody has been presented. The initial structure was prepared from the crystal structure reported by Cho et al. (PDB code: 1N8Z) [13]. From this study the following conclusions can be drawn. Firstly, both secondary structures for the full complex and putative interactions between subdomain IV of ErbB2 and variable domains (V_H and V_L) of Trastuzumab are well conserved

along the molecular dynamics trajectory. Secondly, a hinge move approaching the subdomain II to the Trastuzumab component is revealed by combining the MD trajectory and principal component analysis. This global motion allows the interaction between the dimerization arm of the ECD-ErbB2 subdomain II and subdomain C_H of the antibody. The effect of this interaction on the heterodimerization of ErbB2 and other EGFR receptors is under study in our group. In addition, we have observed some differences between the MD simulation and the x-ray structure attributed to the crystal packing. Thus, the monomer packing in the crystalline cell hinders the hinge move discussed above due to the presence of two other nearby ErbB2/Tzb complexes, thus preventing the interaction between the dimerization arm and the C_H domain of its own Tzb protein. In any case, we expect that these results are useful to identify the underlying interaction mechanism between receptor and antibody, which could help to design new therapeutic antibodies. The observed interaction of the antibody with the dimerization arm could provide us with new clues to design possible modifications of the antibody that may then enhance its therapeutic power. An effective blockade of the dimerization arm would impede the ErbB dimerization and consequently would lead to the interruption of the signaling cascade. The simultaneous effect on both ErbB2 subdomains II and IV exerted by a modified Tzb would be of great interest in the treatment of ErbB2 over-expressed tumors.

4.8 *Additional Information*

Secondary Structure Analysis

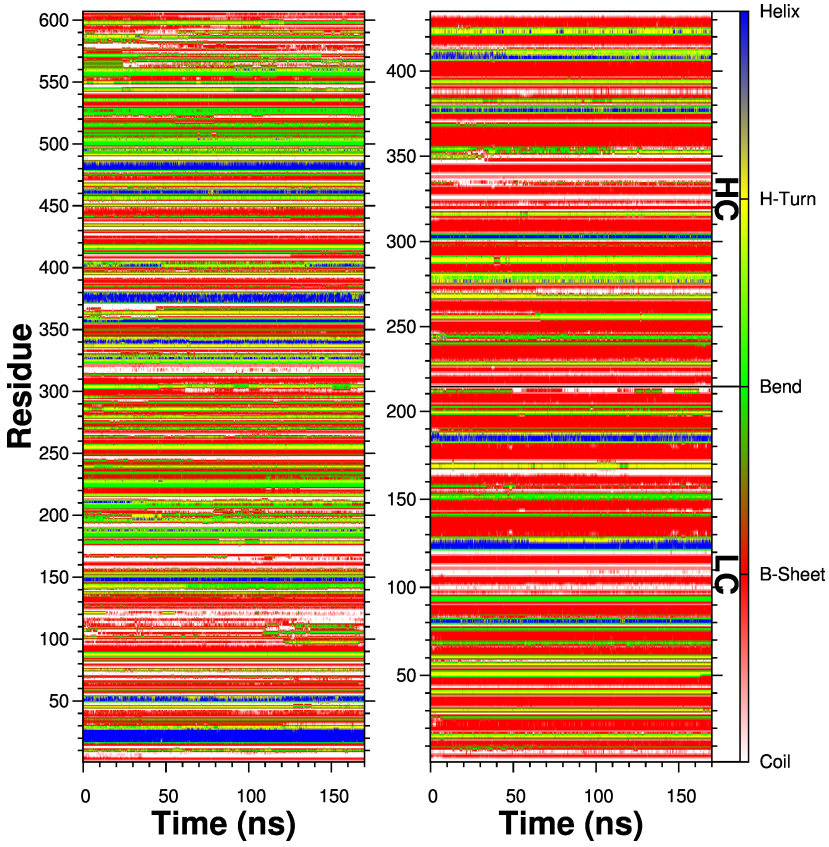


Figure 4.9: Secondary structures along the trajectory of the different domains found in the ErbB2-ECD/TrastuzumabFab complex as calculated by DSSP analysis

Stability Analysis

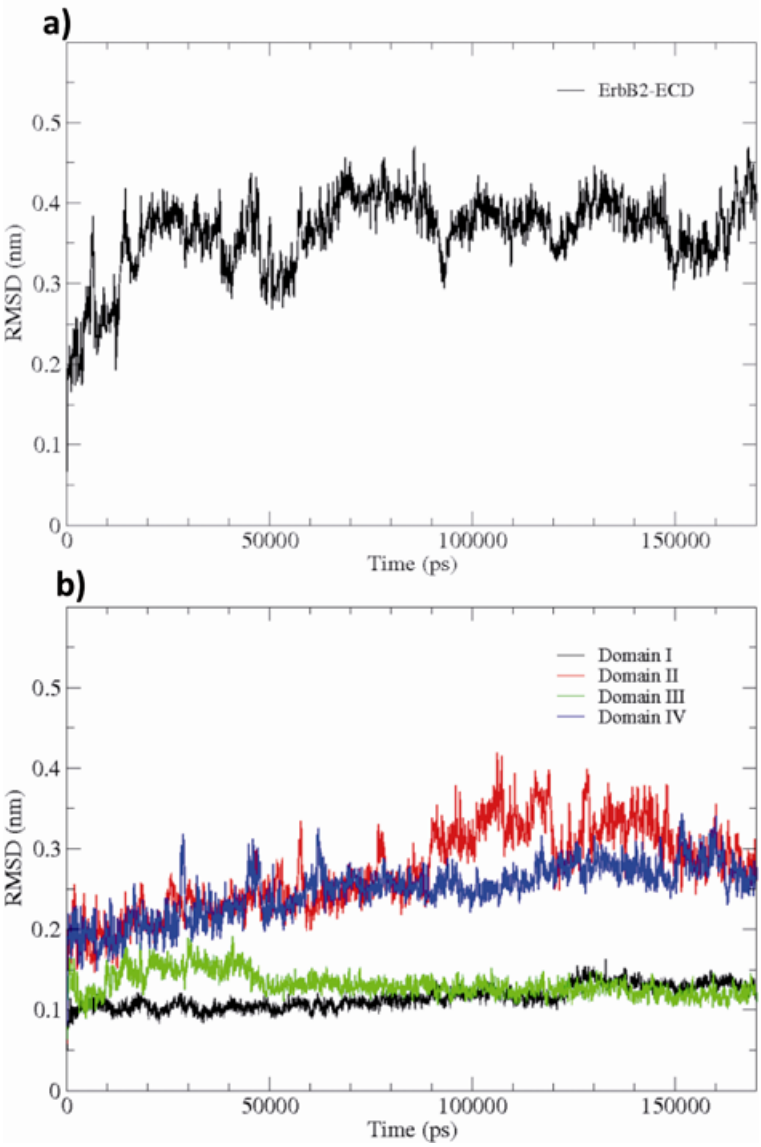


Figure 4.10: **a)** Root mean square deviations (RMSD) as a function of simulation time of the Erbb2ECD receptor backbone atoms in the complex.**b)** RMSD along the MD simulation for the backbone atoms for each of the individual domains defined in the Erbb2ECD protein (Domain I in black, II in red, III in green and IV in blue). The complex Erbb2ECD/Tzbfab crystal structure (PDB code: 1n8z) was used in all cases as reference structure

MM-PBSA to calculate binding energy

The protein-ligand binding free energy for the ErbB2-Trastuzumab complex was calculated using the following equation:

$$\Delta G_{bind} =< G_{complex} > -(< G_{protein} > + < G_{ligand} >) \tag{4.1}$$

where $G_{complex}$, $G_{protein}$ and G_{ligand} are the free energies of the complex (ErbB2-Trastuzumab), the apo-protein (ErbB2) and the ligand (Trastuzumab), respectively. The symbols < and > denote average over snapshots taken from the MD trajectories. We ran 170-ns simulations of the apo-ErbB2 and Trastuzumab isolated systems in solution using the same protocol described in the MD simulation section. In all cases, 100 snapshots were extracted from the last 10 ns of the production runs. For each snapshot the free energy can be evaluated as follows:

$$G_x = E_{MM} + G_{sol} - TS_{MM} \tag{4.2}$$

Thus, the free energy for each component x (i.e, complex, protein and ligand) is evaluated by summing the configurational entropy (S_{MM}), the solvatation free energy (G_{sol}), and gas-phase molecular mechanical energy (E_{MM}). The E_{MM} term is calculated as,

$$E_{MM} = E_{internal} + E_{vdw} + E_{elec} \tag{4.3}$$

Here, $E_{internal}$ is the sum of bond, bend and dihedral energies, E_{vdw} is the van der Waals energy and E_{elec} is the electrostatic energy evaluated in the gas-phase for each component using the OPLS force field with no cut-off.

Solvation free energies may be separated into polar (electrostatic) and nonpolar components,

$$G_{sol} = G_{polar} + G_{nonpolar} \tag{4.4}$$

The polar contribution (G_{polar}) was calculated by solving the finite-difference Poisson-Boltzmann equation using Delphi v5.1 [36, 37]. In the Delphi calculations a grid space of 0.5 Åand a grid size of 291. The OPLS radii and charges were employed for all atoms. The exterior and interior (solute) dielectric constants were set to 80 and 2, respectively. The nonpolar contribution was determined on the basis of accessible surface area (SASA) using the following expression:

$$G_{nonpolar} \approx \gamma SASA + b \tag{4.5}$$

with $\gamma = 0.0054\ kcal/mol$ and $b = 0.92\ kcal/mol$ [38, 39].

The entropy consists of three terms:

$$S = S_{tr} + S_{rot} + S_{vib} \tag{4.6}$$

for the translational, rotational and vibrational entropy, respectively. The expressions for translational and rotational entropies are given for the following expressions according to the reference [40]:

$$S_{tr} = Rln\frac{Ve^{\frac{3}{2}}}{\Lambda_L^3} \tag{4.7}$$

$$S_{rot} = Rln\frac{8\pi^2 e^{\frac{3}{2}}}{\sigma\Lambda_L^x\Lambda_L^y\Lambda_L^z} \tag{4.8}$$

Where V is the volume of the complex, protein or ligand, e is the natural logarithm base, $\Lambda_L = \frac{h^{1/2}}{2\pi m k_B T}$ is the translational thermal de Broglie wavelength, σ is the symmetry number , $\Lambda_L^i = \frac{h^{1/2}}{2\pi I_L^i k_B T}$ is the rotational thermal de Broglie wavelength and I_L^i represents the moments of inertia about the principal components denoted by x, y and z. The vibrational entropy of each system (protein, ligand and complex) was estimated using the quasi-harmonic analysis. This method has been used previously to evaluate the entropy of peptides and proteins [41, 42]. Details of the calculation can be seen in those references. In summary, the entropy contribution to the free energy of binding was using the Schlitter formula [43] by calculating the covariance matrix of fluctuations for frames of each individual system extracted from the MD trajectories.

The binding free energy has been calculated according to the MM-PBSA (Molecular Mechanics-Poisson-Boltzmann Surface Area) method. The calculated binding energies of Trastuzumab ligand to ErbB2 are shown in Table 4.1. The "*in-silico*" binding energy ΔG_{bind} is quite large (-285.0 kcal/mol) in comparison with the reported experimental values between -12.4 and -14.0 kcal/mol [34, 35]. In this sense, Fuentes et al. [12] reported an "*in-silico*" value of $\Delta G_{bind} = -1144.6$ kcal/mol using MM-GBSA (Molecular Mechanics-Generalized-Born Surface Area) approximation without entropic terms. Thus, it seems clear that entropic terms are needed in order to improve the binding energy between the apo-ErbB2 protein and the Trastuzumab ligand.

Table 4.1: Binding free energy components (in kcal/mol) of the ErbB2-Trastuzumab complex calculated from MD trajectory

Energy Term	**ErbB2-Trastuzumab**		**ErbB2**		**Trastuzumab**		**Δ**	
	Mean	σ^c	Mean	σ^c	Mean	σ^c	Mean	σ^c
ΔE_{int}	41811.8	9.5	23064.0	6.6	18808.9	6.2	-61.1	13.1
ΔE_{vdw}	-8061.6	3.8	-4633.4	2.9	-3331.7	2.8	-96.5	5.5
ΔE_{elec}	-65145.8	14.8	-35717.1	10.5	-28367.6	9.5	-1061.1	20.4
ΔE_{MM}	-31395.6	18.0	-17286.5	12.7	-12890.4	11.7	-1218.7	24.9
ΔG_{polar} [a]	-5054.2	5.4	-3586.9	4.8	-2050.3	3.9	582.4	8.2
$\Delta G_{nonpolar}$ [b]	-288.9	1.8	-177.7	1.2	-118.6	1.0	7.4	2.4
$-T\Delta S_{trans}$	18.5	1.9	17.7	1.8	17.2	1.7	16.4	3.1
$-T\Delta S_{rot}$	81.9	8.2	81.0	8.0	80.3	8.2	79.4	14.1
$-T\Delta S_{vib}$	324.7	3.2	297.7	3.0	274.8	2.8	247.8	5.2
ΔG_{bind}	-37163.8	20.9	-21447.5	16.2	-15431.6	15.2	-285.0	30.5

[a] The polar contribution (G_{polar}) was calculated by solving the finite-difference Poisson-Boltzmann equation using Delphi v5.1.
[b] The nonpolar contribution was determined on the basis of accessible surface area (SASA).
[c] Standard error of mean values, $\sigma = standarddeviation/N^{1/2}$

References

[1] Garnier N, Genest D, and Genest M. *Biophys. Chem.*, 58:225–237, 1996.

[2] Samma Soumana O, Aller P, Garnier N, and Genest M. *J. Biomol. Struct. Dyn*, 23:91–100, 2005.

[3] Duneau JP, Genest D, and Genest M. *J. Biomol. Struc. Dyn.*, 13:753–769, 1996.

[4] Garnier N, Crouzy S, and Genest M. *J. Biomol. Struc. Dyn.*, 21:179–199, 2003.

[5] Samma Soumana O, Garnier N, and Genest M. *Eur. Biophys. J.*, 37:851–864, 2008.

[6] Bocharov EV, Mineev KS, Volynsky PE, Ermolyuk YS, Tkach EN, Sobol AG, Chupin VV, Kirpichnikov MP, Efremov RG, and Arseniev AS. *J. Biol. Chem.*, 283:6950–6956, 2008.

[7] Telesco SE and Radhakrishnan R. *Biophys. J.*, 96(6):2321–2334, 2009.

[8] Shih AJ, Telesco SE, Cho SH, Lemmon MA, and Radhakrishnan R. *Biochem. J.*, 436:241–251, 2011.

[9] Telesco SE, Shih AJ, Jia F, and Radhakrishnan R. *Mol. Biosyst.*, 7:2066–2080, 2011.

[10] Wang JN, Feng JN, Yu M, Xu M, Shi M, Zhou T, Yu XD, Shen BF, and Guo N. *Mol. Immunol.*, 40:963–969, 2003.

[11] Hu P, Feng J, Zhou T, Wang J, Jing B, Yu M, Hu M, Zhang X, Shen B, and Guo N. *J. Cell Physiol.*, 205:335–343, 2005.

[12] Fuentes G, Scaltriti M, Baselga J, and Verma C. *Breast Cancer Res.*, 13(3):R54, 2011.

[13] Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, and Leahy DJ. *Nature*, 421(6924):756–760, 2003.

[14] Tripos International. Sybyl 8.0. St. Louis, MO.

[15] Dayhoff M, Schwartz R, and Orcutt B. *Atlas of protein sequence and structure*. National biomedical research foundation, Washington DC, 1978.

[16] Hess B, Kutzner C, van der Spoel D, and Lindahl E. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[17] Jorgensen WL and Tirado-Rives J. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.

[18] Jorgensen WL and Tirado-Rives J. *Proc. Natl. Acad. Sci. USA*, 102(19):6665–6670, 2005.

[19] Kaminski GA, Friesner RA, Tirado-Rives J, and Jorgensen WL. *J. Phys. Chem. B*, 105:6474–6487, 2001.

[20] Berendsen H, Postma J, van Gunsteren W, and Hermans J. *Intermolecular forces*, chapter Interaction models for water in realtion of protein hydration, pages 331–342. Reidel, Dordrecht, 1981.

[21] Darden T, York D, and Pedersen L. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[22] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[23] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[24] Hess B, Bekker H, Berendsen HJC, and Fraaije JGEM. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[25] Humphrey W, Dalke A, and Schulten K. *J. Mol. Graph.*, 14(1):33–38, 1996.

[26] Amadei A, Linssen ABM, and Berendsen HJC. *Proteins Struct. Funct. Bioinforma.*, 17(4):412–425, 1993.

[27] Amadei A, Linssen AB, de Groot BL, van Aalten DM, and Berendsen HJ. *J. Biomol. Struct. Dyn.*, 13(4):615–625, 1996.

[28] Kabsch W and Sander C. *Biopolymers*, 22(12):2577–2637, 1983.

[29] Lou H and Cukier RI. *J. Phys. Chem. B*, 110(12796-12808), 2006.

[30] Hunenberger PH, Mark AE, and Van Gunsteren WF. *J. Mol. Biol.*, 252:492–503, 1995.

[31] Kaiser HF. *Educ. Psychol. Meas.*, 20:141–151, 1960.

[32] JF Franco-Gonzalez, J Ramos, VL Cruz, and J Martínez-Salazar. *J. Mol. Model.*, 19(3):1227–1236, 2013.

[33] Sobolev V, Eyal E, Gerzon S, Potapov V, Babor M, Prilusky J, and Edelman M. *Nucl. Acids. Res.*, 33:W39–W43, 2005.

[34] Troise F, Cafaro V, Giancola C, D'Alessio G, and Lorenzo C. *FEBS J.*, 275:4967–4979, 2008.

[35] Bostrom J, Haber L, Koenig P, Kelley RF, and Fuh G. *PLOS*, 6(4):e17887, 2011.

[36] Rochhia W, Alexov E, and Honig B. *J. Phys. Chem. B*, 105:6507–6514, 2001.

[37] Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M, and Alexov E. *BMC, Biophys*, 4(1):9, 2012.

[38] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, and Wang W. *Acc. Chem. Res.*, 33(12):889–897, 2000.

[39] Sitkoff D, Sharp K, and Honig B. *J. Phys. Chem. B*, 98:1978–1988, 1994.

[40] Irudayam SJ and Henchman RH. *J. Phys. Chem. B*, 113:5871–5884, 2009.

[41] Pöhlmann T, Böckmann RA, Grubmüller H, Uchanska-Ziegler B, Ziegler A, and Alexiev U. *J. Biol. Chem.*, 279:28197–28201, 2004.

[42] Basdevant N, Weinstein H, and Ceruso. *J. Am. Chem. Soc.*, 128(39):12766–12777, 2006.

[43] Schlitter J. *Chem. Phys. Lett.*, 215:617–621, 1993.

Chapter 5

Mapping from Atomistic to Coarse–Grained Resolution

This chapter is based on the submitted manuscript *"Comparing Atomistic and Martini Coarse-Grained Models on Protein-Protein Interactions applied to the ErbB2/Trastuzumab Complex"* by JF Franco-Gonzalez, J Ramos and VL Cruz

Abstract

The coarse grained Martini force field has been compared with the atomistic OPLS representation. We tested the Martini model considering variants of the following molecular dynamics settings: the non-bonded interactions methods to calculate the electrostatic interactions, the value of the neighbor lists cut-off radius (r_{list}) and the Elastic Network method. The ectodomain-ErbB2/Trastuzumab-Fab complex, which is of relevant importance in anticancer therapy, was selected to analyze the protein-protein interactions. The results show that when used in molecular dynamics simulations domElNeDyn models, PME and an r_{list} 1.4 nm, are comparable to the atomistic protein models. The results support the value of the Martini force field in the study of protein-protein interactions and towards protein structure prediction.

5.1 Introduction

The coarse graining simulation techniques emerge as a useful paradigm to perform simulations at the microsecond time scale on protein complexes of large size. Particularly, the Martini force field has been established as a suitable model for the coarse-grained (CG) description of biological systems [1, 2]. Four atoms per CG particle are mapped in Martini, which offers parameters for several types of molecules including, essentially, lipids, carbohydrates and proteins. However, the appropriate usage of Martini on protein-protein interactions is still a matter of research [3–5]. For example, Stark et al. suggested a scaling down of the Van de Waals parameters that describe the interactions between protein pseudoatoms to reproduce experimental data on the thermodynamics of protein-protein interactions in aqueous solution [5]. Another aspect investigated by some authors focused on the use of an elastic network (EN) protocol named ElNeDyn [3], which is a necessary addition to the Martini force field when studying protein-protein association processes. An elastic network by domains (domElNeDyn) is a variation of the original EN protocol that permits more flexibility of the interdomain regions of the protein while maintaining, at the same time, the overall shape of the protein [6]. Several authors reported the effect of different parameters of the domElNeDyn protocol when simulating different protein complexes [6, 7]. A suitable knowledge of the protein domain composition and stability is required to get reasonable results [8]. Hence, although the Martini force field has improved parameter sets [9], it still needs further work to get better in the protein-protein interactions and specially in structure predictions [4, 10, 11].

Therefore, there is considerably interest to understand better the behavior of the Martini force field regarding protein complex systems. In this work we explore the effect of three simulation parameters to be combined with the Martini force field, namely, the non-bonded methods to calculate electrostatic interactions, the value of the neighbor lists cut-off radius and the Elastic Network method. To examine the first one we perform simulations with the Shift model to treat electrostatic interactions, which is the original choice of the Martini parameterization [2], and Particle Mesh Ewald (PME) [12, 13]. Additionally, we explore the effect of increasing the neighbor lists cut-off radius from the original value of 1.2 nm [1, 2, 5, 7, 9, 14] to 1.4 nm. The third variable we have taken into account were both elastic networks protocols: ElNeDyn and domElNeDyn.

In this work, we have chosen the ECD-ErbB2/Trastuzumab-Fab complex to analyze the performance of CG models when compared to all-atomistic simulations (AA) of the same system. This protein complex offers high flexibility, very well defined domains, stable protein-protein contacts along the molecular dynamics and large movements that involve interfaces with additional protein-protein interactions [15, 16]. We have carried out a structural comparison between AA and CG simulations at two different time windows. On one hand, the AA and

CG trajectories are matched considering the time window of 500 ns, which is the period used for the atomistic simulations. On the other hand, a larger time window of several microseconds of CG dynamics was also used for a comparative analysis as well as to assess the stability of the CG models at those large time scales.

We used the clustering analysis proposed by Daura and et al. [17] to get representative structures from the different trajectories in order to facilitate the comparative study [18].

Specifically, the epidermal growth factor receptors (EGFRs) are a family of membrane proteins engaged in a number of biological processes such as cellular proliferation. The over expression of some of these receptors at the cell surface has been associated with cancer [19]. Particularly, ErbB2 over expression on breast cancer cells has been identified as a negative factor affecting cancer progression and final survival time [20, 21]. The monoclonal antibody Trastuzumab (Trastuzumab) has been designed to target an extracellular epitope of the ErbB2 receptor [22–24], being the first specific anti-ErbB2 treatment approved by the Food and Drug Administration (FDA). However, the molecular mechanism of action of Trastuzumab is not well understood [25–27]. Some mechanisms consider that Trastuzumab inhibits activation of ErbB2 disrupting the cleavage of the ectodomain region caused by the interaction between ErbB2 and proteases [28]. In fact, Cho and et al. [29] elucidated a crystallographic structure of the ECD-ErbB2/Trastuzumab-Fab complex, where Trastuzumab binds to subdomain IV of ECD-ErbB2.

The ErbB2 ectodomain consists of four very well defined subdomains classified according to leucine-rich domains (I and III) and cystein-rich domains (II and IV) [28–35]. This receptor is distinguished by an extended conformation in absence of the ligand unlike the rest of family receptors [30]. This kind of conformation makes it possible to form dimers with other receptors through an exposed dimerization arm that is located in subdomain II. Regarding the dimerization as a therapeutic target, a monoclonal antibody known as pertuzumab was developed to bind to the epitope in the subdomain II [36]. The ErbB2-Trastuzumab complex has been studied by molecular dynamics simulation (MD) at both resolution scales, namely, AA and CG. These simulations have revealed an additional interaction between the antibody constant domain and the dimerization arm [6, 15, 16, 37]. In this chapter a direct comparison between both models is performed.

The chapter is organized as follows: The Computational Methods section describes the different tools and settings used to perform the simulations and analysis of the atomistic and coarse-grained models. The Results section presents in the first two subsections an analysis of the conformational flexibility exhibited by the atomistic and coarse-grained simulations. In the two following subsections the clustering tool is used to select the representative structures to perform

the comparative analysis. An additional Discussion section is provided to give a detailed description of the ErbB2-Trastuzumab complex interface. Finally, the Conclusions section summarizes the most relevant results obtained from the present work.

5.2 Computational Methods

General Settings

MD simulations of ECD-ErbB2/Trastuzumab-Fab complex were carried out starting from the crystal structure available in the Protein Data Bank server as PDB code: 1N8Z [29]. The missing residues in the loops 102-110, 303-305, 361-364 and 581-590 have been modeled based on homologous sequences as was reported in reference [16] for the same system. Trastuzumab-Fab component consists of two protein chains: light chain (LC) and heavy chain (HC), each of one with their variable and constant domains labeled as V_L and C_L for LC and V_H and C_H for HC. For the ECD-ErbB2 protein, four different domains are described, namely, domains I, II, III and IV or similarly leucine-rich domains 1 and 2 (L1 and L2) and cystein-rich domains 1 and 2 (CR1 and CR2). We assumed that the pKa of the individual amino acid residues at physiological pH does not change in the protein receptor. Thus, histidine (H) residues remained neutral; lysine (L) and arginine (R) were protonated and aspartic (D) and glutamic (E) acids were deprotonated. The resulting total charge for the complex was -10 e units. In all setups (including the coarse grained simulations) 10 Na^+ counter ions were added to yield an electrically neutral system. Periodic boundary conditions were applied along the three dimensions. The setups were relaxed along 2500 steps using the steepest descent minimization algorithm. The GROMACS package version 4.6.3 [38] was used for all simulations. The pressure and temperature were kept constant at 1 bar and 300 K, respectively, using the Berendsen coupling algorithm [39] by coupling the proteins (ECD-ErbB2 and Trastuzumab-Fab), solvent and ions independently with isotropic scaling for the pressure (1 bar) and a compressibility factor of $4.5x10^{-5}\text{bar}^{-1}$.

Atomistic Simulation

The OPLS [40] force field for the protein and SPC [41] water model for the solvent were used in this simulation. The system was solvated by 60717 water molecules. The temperature and pressure were kept constant with the time constants $\tau_T = 0.2ps$ and $\tau_P = 1ps$. The Particle Mesh Ewald method (PME) [12, 13] was used for the long-range electrostatic interactions with a maximum grid spacing of 0.25 nm and using fourth-order (cubic) interpolation for the fast Fourier transforms. A short-range electrostatic cut-off radius of 1.0 nm, a Van der Waals cut-off

radius of 1.0 nm and a short-range neighbor list cutoff of 1.0 nm were used. Bond lengths were constrained using the LINCS algorithm [42] for all bonds in the protein. The system was equilibrated for 2 ns at constant pressure and temperature (NPT ensemble) with position restraint for all protein atoms with a force of 1000 $kJmol^{-1}nm^{-2}$. Production dynamics were performed at NPT releasing all constraints on the heavy atoms during 500 ns and accumulating the trajectory frames every 10 ps without any restraints.

Coarse-Grained Simulations

The simulations were performed with MARTINI CG 2.1 [2]. The coarse-grained representation (topologies and parameters) of the ECD-ErbB2/Trastuzumab-Fab complex was generated from a minimized atomistic structure. Standard MARTINI CG water beads were used to model the solvent [43] (a total of 38986 beads). The setups were energy-minimized and the solvent and ions were relaxed with position restraints (1000 $kJmol^{-1}nm^{-2}$) applied to the backbone beads. The relaxation with position restraints was performed in two steps, firstly, a NPT protocol with a time step of 2 fs for 1 ns and, secondly, a NPT protocol with a time step of 20 fs for 10 ns. The temperature and pressure were kept constant with time constants $\tau_T = 1ps$ and $\tau_P = 2ps$. The short-range electrostatic and Van der Waals cut-off radius were 1.2 nm and shifted from 0.9 nm for the Lennard-Jones potential and from 0.0 for the electrostatic potential. The neighbor lists cut-off radius used was 1.2 nm and 1.4 nm (updated every 10 steps). To calculate the electrostatic interactions (EIs), Shift and PME non-bonded interactions methods were taken into account. As implemented in GROMACS, the Van der Waals methods considered were Shift and Cut-Off respectively. For Martini CG simulation it has been shown that the simulation time typically should be multiplied by a factor of 4 to roughly account for the increase in diffusion observed for CG water beads [1, 43]. Thus, in this work the simulation time is multiplied by 4 resulting in an "effective" time. All production simulations were run using NPT with 20 fs time step for 10 μs (40 μs effective time).Two replicas of each one were run with different seed number. Every CG protein model was built using the martinize.py script downloaded from http://cgmartini.nl

Elastic Network

The ElNeDyn [3] method was used to assign the elastic network to each protein (an elastic network for ErbB2 protein and an elastic network for each chain of Trastuzumab-Fab) in two ways: for the entire proteins without distinction between domains (ElNeDyn) and by domains in each protein (domElNeDyn). In ElNeDyn simulations, an elastic network (EN) is built on the backbone beads of the MARTINI protein model to maintain the initial tertiary structure as is

Table 5.1: Simulations performed in the present chapter

FF[a]	EI Method	EN	r_{list}(nm)	# Rep.[b]	Time (μs)	Label
OPLS	n.a	n.a	1.0	1	0.5	**AA**
Martini	PME	Elnedyn	1.2	2	10	**P1.2**
			1.4	2	10	**P1.4**
		domElnedyn	1.2	2	10	**Pd1.2**
			1.4	2	10	**Pd1.4**
	Shift	Elnedyn	1.2	2	10	**S1.2**
			1.4	2	10	**S1.4**
		domElnedyn	1.2	2	10	**Sd1.2**
			1.4	2	10	**Sd1.4**

[a] FF. *Force Field*.
[b] Rep. *Replica*.

described by Periole et al [3]. It is well-known that MARTINI protein models in water need an EN to further constrain the protein close to a particular conformation, e.g. native state [3, 7, 14]. On the other hand, based on the knowledge of the EGFR receptor, the ECD-ErbB2 [29, 44] domains are well defined, therefore it is possible to assign domElNeDyn. We have compared the domains defined in this way with the assignment by the DomFOLD method [8] via its online server [45]. We have found that there are not significant differences between these two domain definitions. Additionally, the work by Siuda and et al. concluded that " domElNeDyn simulations do not depend significantly on the exact domain assignment" [7].

The parameters controlling the elastic networks were a cutoff radius $R_c = 0.8nm$ and a spring constant $K_{spring} = 500kJmol^{-1}nm^{-2}$. These are the values to reach the best overlap with the atomistic simulations, as were optimized by Periole et al [3]. Additionally, these parameters values have been chosen according to the previous reports published elsewhere [6, 7]. A complete list of simulations and the number of replicas of each one are shown in Table 5.1.

GPU simulations

PME dynamics simulations were run on local workstations equipped with 4 Nvidia GPUs each, called Metrocubo and bought to Acellera Ltd. Metrocubo. This is a highly compact hardware-software solution for running GPU accelerated molecular dynamics simulations that can reach supercomputing performance.

Conformational Stability Analysis

Root mean square deviation (RMSD) has been used to examine the conformational stability of the complex and the root mean square fluctuations (RMSF) was calculated on individual residues of the complete trajectory. Both RMSD and

RMSF were determined on the Cα atoms for the AA simulation or the backbone beads (BB beads) for the CG simulations. The minimized atomistic structure was used as the reference to compare RMSD as well as RMSF.

Principal Component and Residue-Mean-Square Inner Product ($RMSIP$) analysis

Principal component analysis (PCA) calculates the dominant modes in the motion of the molecule along the MD simulation trajectory. Thus, the configurational space is reduced, containing few relevant collective degrees of freedom in which long range fluctuation can be studied [18, 46]. PCA diagonalizes the covariance matrix of the atom fluctuations from their average trajectory. The resulting eigenvectors correspond to the dominant low frequency motions. We have performed a PCA analysis in order to identify these lowest frequency motions occurring in the ECD-ErbB2/Trastuzumab-Fab complex. The g_covar and g_anaeig tools in the GROMACS package were used to perform the PCA analysis. The PCA diagonalization yields the principal directions of the large-amplitude concerted motions that characterize the essential subspace of the internal dynamics of the protein. The RMSIP between the essential subspaces of both AA and CG simulated systems has been used to assess their dynamical similarity. The RMSIP was calculated on the first ten eigenvectors along the complete MD trajectory using the Bio3D package [47].

Clustering Analysis

Due to the complexity and big size of the data produced in this work, we have done a clustering analysis on the trajectories to be able to compare between AA and CG conformations obtained with the different methodologies. Each frame in the AA trajectory was mapped to a CG representation. Clustering based on pairwise RMSD was performed using a previously reported clustering algorithm, namely, "gromos" implemented in GROMACS (g_cluster) [17]. In this method an RMSD cut-off should be selected to find two neighbor structures. RMSD cut-offs were selected using a pairwise frame to frame RMSD as explained by Karplus et al. [48], McCammon et al. [49] and Schulten et al. [50]. The distribution has been calculated using the positions of the Cα atoms for the AA simulation and BB beads for the CG simulations. The final cutoff values selected are shown in Table 5.2. The atomistic structures were mapped to the CG representation with the martinize.py tool available in the Martini site (http://md.chem.rug.nl/). In this way a more direct comparison between AA and CG clusters can be done.

5.3 Results

Sections 5.3.1 and 5.3.2 analyze the molecular flexibility in two different time windows. The first one corresponds to the 500 ns interval calculated with the atomistic model (from now on atomistic simulation time window). The second one is extended to several μ-seconds accessible to the CG models (microsecond time window). Sections 5.3.3 and 5.3.4 show the results obtained after clustering the trajectories in order to compare the conformations produced by both AA and CG simulations.

5.3.1 Conformational Flexibility at Atomistic Simulation Time Window

The RMSD and RMSF analysis were done over the "effective" time (see Computational Methods) on CG simulations for the first 500 ns, in order to compare with the AA simulation (See Figures 5.1 and 5.2). In all cases the minimized atomistic structure mapped on CG representation was used as the reference structure. The AA trajectory reaches a relative stability at around 1 nm, as can be seen in Figure 5.1. This relatively large RMSD value is due to the departure of the protein complex geometry from the initial structure containing the crystallographic arrangement. The RMSF profile for the AA simulations shows the largest fluctuations on both group residues 100-110 and 480-490, as it is depicted in Figure 5.2 (subdomains I and IV, respectively). This high flexibility can be explained because these loops are absent in the crystallographic structure and they were built by homology, as reported in [51]. Both the ECD-ErbB2 and Trastuzumab-Fab secondary structures are well conserved along the AA trajectory as can be seen in the Define Secondary Structure of Proteins (DSSP) analysis of the protein (See Figure 5.12 of the Additional Information). These results support the application of the Martini CG model along with the elastic network, where secondary structure has to be preserved along the simulation.

The conformational flexibility in the CG simulation is larger than in the AA one (see Table 5.4). However, when PME with ElNeDyn protocol is used the RMSD profile is lower than or equal to 1.0 nm, whereas with ElNeDyn by domains there are RMSD values higher than 1.0 nm, as shown in Figure 5.1. On the other hand, if one compares CG and AA simulations by RMSF, it is possible to see that, in general, there is a good agreement between them. Regarding only the CG simulations, it is clear that the Shift based methods produce larger RMSF fluctuations than PME. This is particularly evident at shorter r_{list} values.

ElNeDyn protocols have similar fluctuations to the AA simulations while the domElNeDyn presents more differences respect to the AA profiles. In particular, simulations labeled with Pd1.4 and S1.4 show the most similar RMSF profiles to the AA production, as is shown in the Figure 5.2. In summary, the RMSD

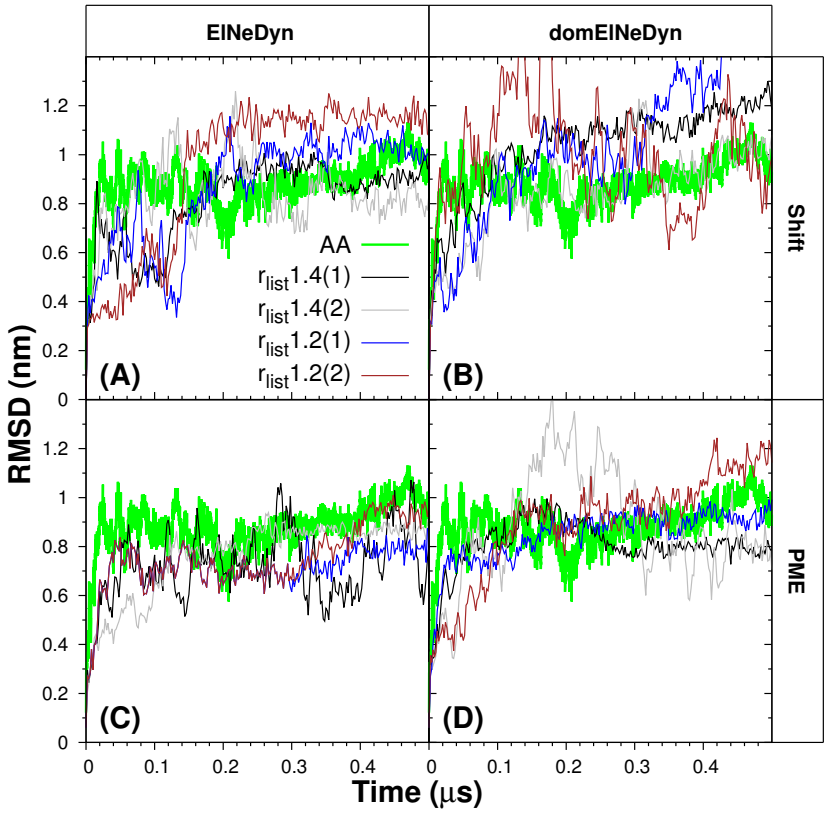


Figure 5.1: RMSD Comparison between CG and AA simulation at the atomistic simulation time window. Every CG simulation was plotted comparing with the different r_{list} and with the corresponding replica: (A) Shift-ElNeDyn, (B) Shift-domElNeDyn, (C) PME-ElNeDyn and (D) PME-domElNeDyn.

and RMSF values show a systematic behavior depending on the choice of Elastic Network protocols, cutoff radius r_{list} and non-bonded electrostatic methods. In general, it can be observed more rigidity with ElNeDyn than with domElNeDyn. Simulations performed with r_{list} equal to 1.4nm are closer to AA values (see Figures 5.1 and 5.2). In order to compare the dominant modes extracted from the different MD simulations, a PCA analysis followed by a RMSIP evaluation was carried out following the methodology described in the Computational Methods section. The RMSIP values, in all cases, are greater than 0.5. Therefore, a good overlapping should be expected since the CG simulations show both flexibility and conformational stability similar to the AA, as was suggested in reference [7]. The RMSIP seems to behave similarly independently of the different CG protocols, because no significant differences between the different CG models are observed (see Table 5.6). It might be safe to suppose that the CG simulations sample a similar conformational phase space than the AA simulation. It can be

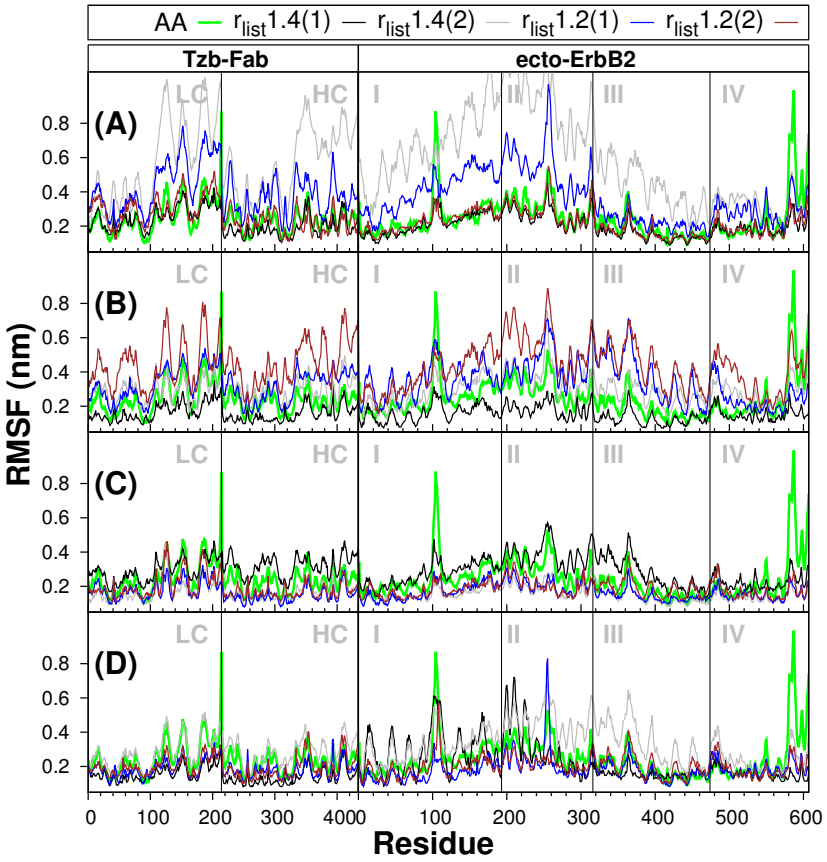


Figure 5.2: RMSF Comparison between CG and AA simulations at the atomistic simulation time window. Every CG simulation was plotted comparing with the different r_{list} and with the corresponding replica: (A) Shift-ElNeDyn, (B) Shift-domElNeDyn, (C) PME-ElNeDyn and (D) PME-domElNeDyn.

assumed, from the above results, that 500 ns is an acceptable production time to find a stable conformation in this protein-protein complex.

5.3.2 Flexibility and Dynamics at Microsecond Time-Scale

An extension of the CG simulations to μ-second timescales is interesting to further explore the conformational space and compare with the AA results at shorter time scales. We have assumed that 500 ns in the AA simulations are sufficient to reach an equilibrated protein-protein complex (Figures 5.1 and 5.2). The RMSD gives a measure of the stability along the simulation runs at the microsencond scale. In Figure 5.3 the RMSD results for CG simulations are shown. In all cases, the minimized atomistic structure mapped on the CG model was used as a reference structure. In addition to this, RMSFs were calculated on each residue for both

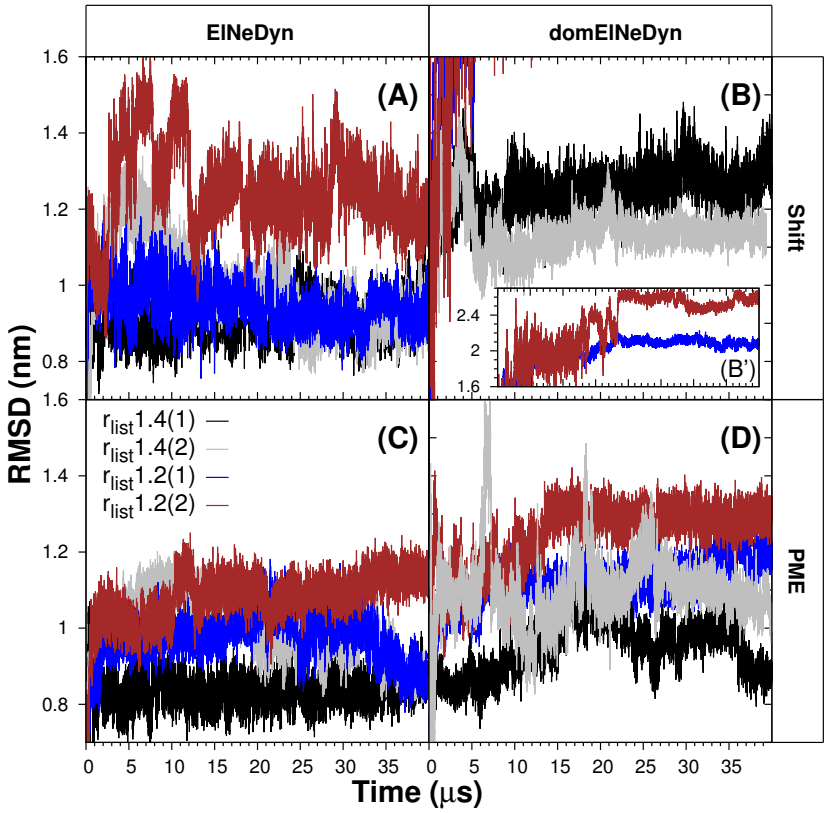


Figure 5.3: RMSD for the BB particles as a function of effective time in the Martini CG simulations. The minimized atomistic structure converted to CG was used as reference. (A) Shift-ElNeDyn methods, (B) Shift-domElNeDyn [the inset (B) shows an extension on the vertical axis], (C) PME-ElNeDyn methods, and (D) PME-domElNeDyn

AA and CG simulations as can be seen in Figure 5.13. The complete trajectory was used to calculate the RMSF.

In general, the CG RMSD profiles show larger fluctuation than the AA RMSD ones, as can be seen in Figure 5.3 and Table 5.5 . This is a standard behavior on Martini protein models in water and that is the reason why it is necessary to add EN to the models [3]. ElNeDyn shows RMSD values higher than domElNeDyn when the r_{list} and EI method are the same, as can be seen in Figures 5.3A- 5.3B and Figures 5.3C-5.3D. At equal EN protocol and r_{list}, the RMSD values are higher for Shift than for PME (Compare Figures 5.3A-5.3C and Figures 5.3B-5.3D). Finally, the RMSD values are always higher for a $r_{list} = 1.2nm$ than for a $r_{list} = 1.4nm$ under the same EN protocol and EI method (See Figures 5.3A-D).

In Figure 5.13 the RMSF values are presented for the different CG models compared to the atomistic simulation. Shift-ElNeDyn and PME-domElNeDyn show a good agreement with the atomistic fluctuations (see Figure 5.13B and

Table 5.2: The RMSD cutoff selection for the different methodologies

Method	$RMSD_{cutoff}$ (nm)	# Clusters
AA	0.35	12
P1.2	0.26	16
P1.4	0.39	8
Pd1.2	0.45	8
Pd1.4	0.73	6
S1.2	0.38	15
S1.4	0.51	7
Sd1.2	0.63	26
Sd1.4	0.44	15

5.13C). The worse match to AA fluctuations corresponds to the Shift-domElNeDyn method which, additionally, shows a replica with very large fluctuations. The domElNeDyn simulations present more variability between the two replicas in contrast with the uniformity of the ElNeDyn replicas due to the additional flexibility shown by the interdomain freedom in the former method. In general, the subdomain IV C-terminal fragment is less flexible in the CG representations that in the AA simulations. This fragment along with the homology built loop in subdomain I are highly flexible in the AA system and not well captured by the CG models.

Finally, the RMSIP analysis performed between the μ-second CG trajectories and the AA simulation shows very similar results to that reported for the 500 ns interval except for the Shift-domElNeDyn method, as can be observed by comparing Tables 5.6 and 5.6.

5.3.3 Cluster Analysis to Find out Structural Similarity

The clustering analysis is performed to facilitate the selection of average conformations from the simulations and further comparison between AA and CG trajectories.

In order to select a cutoff value for generate the clusters, a pairwise frame to frame RMSD distribution along the AA and the CG trajectories are calculated and shown in Figure 5.4. The RMSD cutoff used to generate the clusters for every method considered in the present work was selected according to the criteria described in the Computanional Methods section and reported by Schulten et al [50]. Table 5.2 presents the values chosen for every model.

Let us keep in mind that the AA structures were mapped on the CG representation to facilitate a direct comparison with the CG models. A total of 12 clusters were found applying a RMSD criterion of 0.35 nm to discriminate conformations in the AA simulation taking into account the full 500 ns trajectory

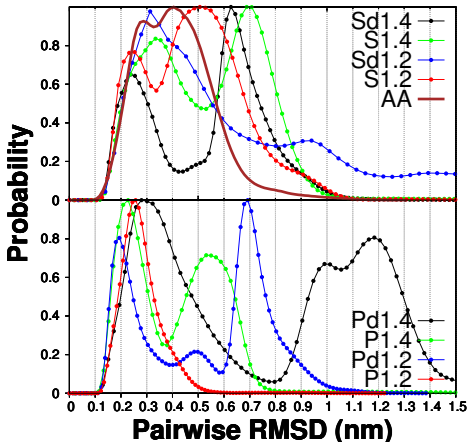


Figure 5.4: Pairwise Frame to Frame RMSD distribution calculated on the Cα atoms for the AA simulation and BB beads for the CG simulations. *Top*: Shift methods. *Bottom*: PME methods.

(See Table 5.2). The first and second clusters represent 70.5 and 13.1 % of the trajectory, as is shown in Figure 5.14. These clusters amount a total of 83.6 % of the production dynamics and therefore, we have not taken into account the remaining clusters in the subsequent analysis. Representative structures corresponding to average geometries for both clusters are shown in Figure 5.5. The representative structures were depicted emphasizing on the protein interaction between the ErbB2 and the antibody Fab.

We have performed the same analysis for each CG trajectory. Similarly to AA simulations, we have taken into account only the two more representative structures of the cluster analysis. Clusters 1 and 2 from the AA simulation were used as reference. The percentage distribution for the RMSD values calculated on the frames corresponding to each CG trajectory is represented in Figure 5.6. The Pd1.4 model gives a good match with the structure represented by cluster 2 of the AA simulation and slightly worse with cluster 1, as shown in Figure 5.6. In general, Pd1.4 shows the higher percentage values, giving conformations with small RMSD (RMSD less than 0.7 nm) for both clusters. However, despite the fact that Pd1.4 shows good percentage values they have a high variability around the mean value (see the error bars in Figure 5.6). This points out that the two different production runs, on average, compare satisfactorily well with the AA system.

5.3.4 Cluster Analysis to Study the Protein Complex Interface

AA simulations

A more detailed description of the protein-protein interaction on each method can be obtained by the calculation of distance matrices between the BB particles of the ECD-ErbB2 and the Trastuzumab-Fab protein moieties. The corresponding distance matrices for the two clusters derived from the AA simulation are shown

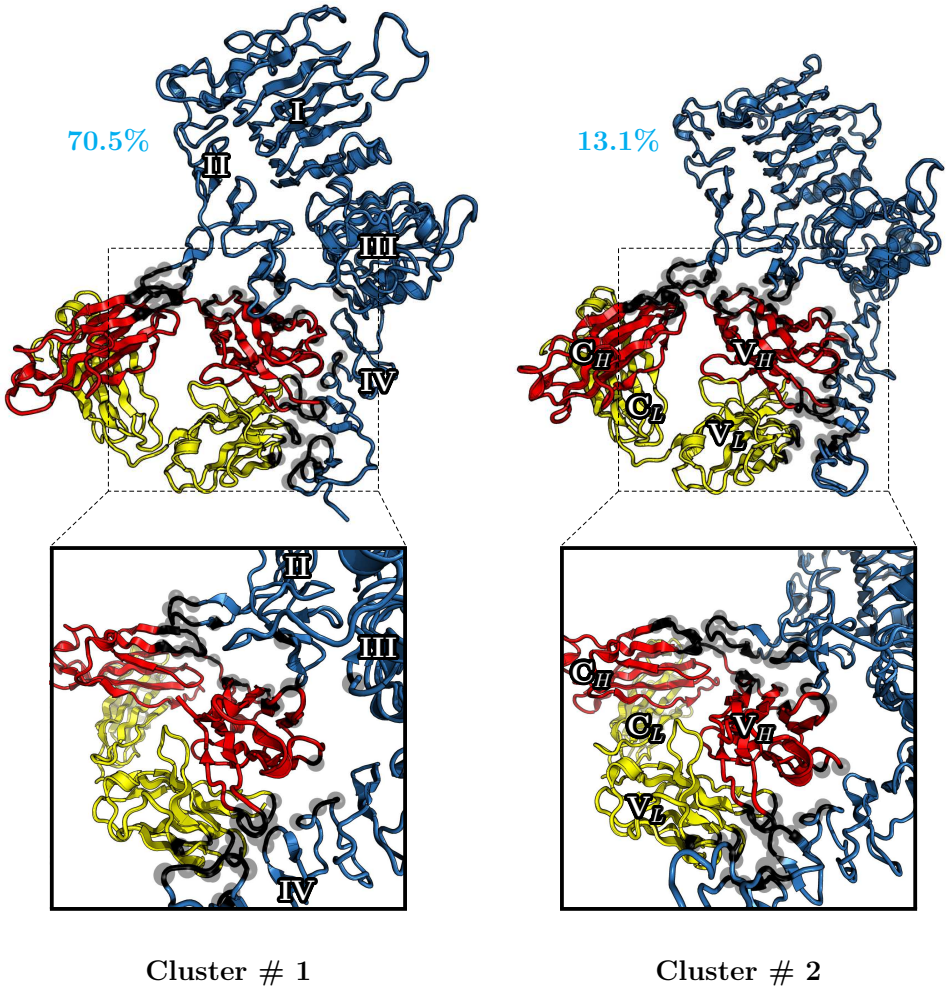


Figure 5.5: Structural comparison between both atomistic clusters 1 (*left*) and 2 (*right*). They are represented by cartoons in which blue is for ECD-ErbB2 while Trastuzumab-Fab is red and yellow for the HC and LC respectively. The inset boxes show a zoom on the interfaces involved in the protein complex interaction. The RMSD between these structures is 0.45 nm

in Figure 5.7. The matrix shows the pattern corresponding to the different interaction interfaces involved in the ECD-ErbB2/Trastuzumab-Fab interaction.

In addition to this, residue pair lists involved in these distance matrix pattern between the proteins in the ECD-ErbB2/Trastuzumab-Fab complex have been calculated and compiled in Tables 5.8-5.12 in the Additional Information section. The pair list corresponds to the pair of residues which BB particles are within a cutoff distance of 0.8 nm. According to this analysis a total of 5 interfaces

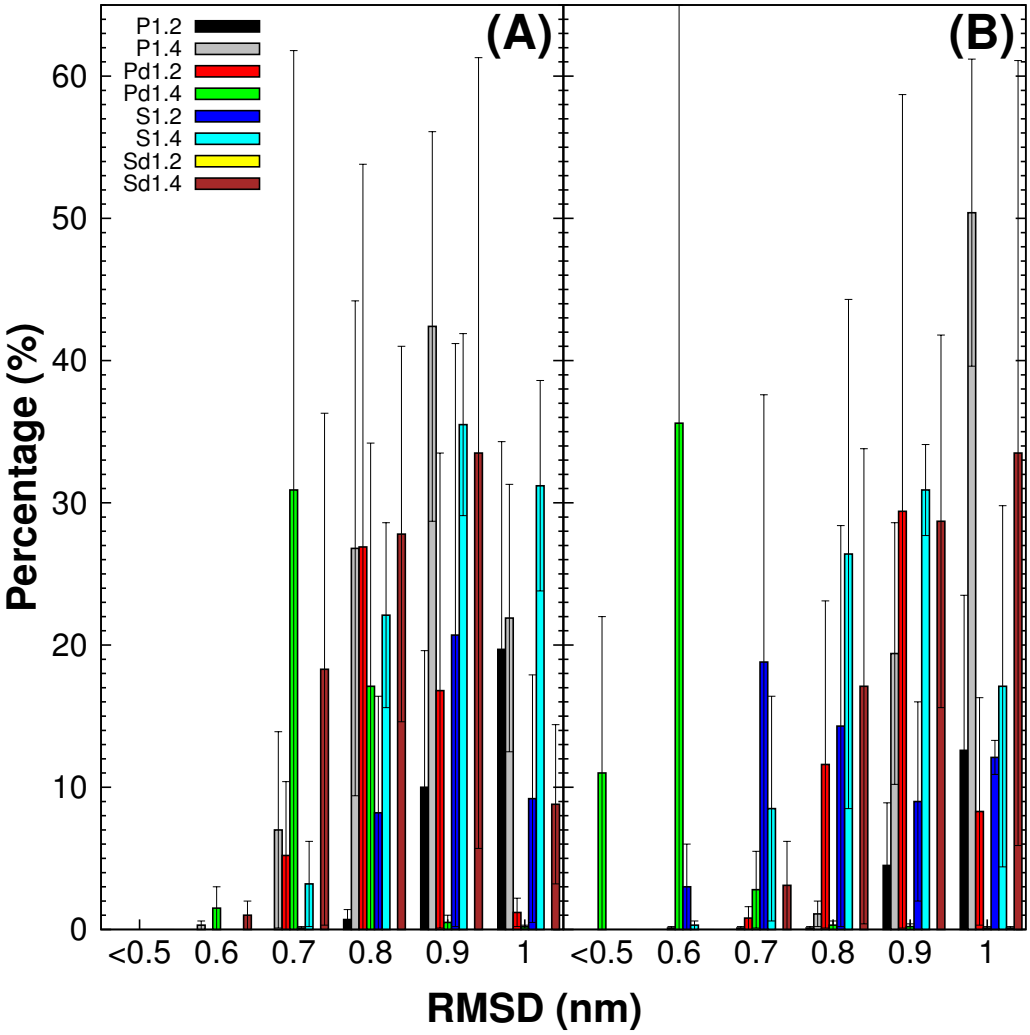


Figure 5.6: RMSD distributions corresponding to each Martini method. Clusters 1 **(A)** and 2 **(B)** from the AA simulation were used as reference. AA structures were mapped to CG structures using the *martinize.py* script. The percentage of structures of CG simulations with different RMSD were calculated. RMSD$\leq$0.5 nm (label $\leq$0.5), 0.5<RMSD$\leq$0.6 nm (label 0.6), 0.6<RMSD$\leq$0.7 (label 0.7), 0.7<RMSD$\leq$0.8 (label 0.8), 0.8<RMSD$\leq$0.9 (label 0.9) and 0.9<RMSD$\leq$1.0 (label 1.0).

were found, namely, subdomains V_L and IV (V_L-d4), subdomains V_H and IV (V_H-d4), subdomains V_H and III (V_H-d3), subdomains V_H and II (V_H-d2) and subdomains C_H and II (C_H-d2).

The reported crystallographic contacts are kept in the structures depicted in the clusters 1 and 2 as can be seen in Figure 5.7 (see below in the Discussion section). The interfaces C_H-d2 and V_H-d2 present the largest differences between

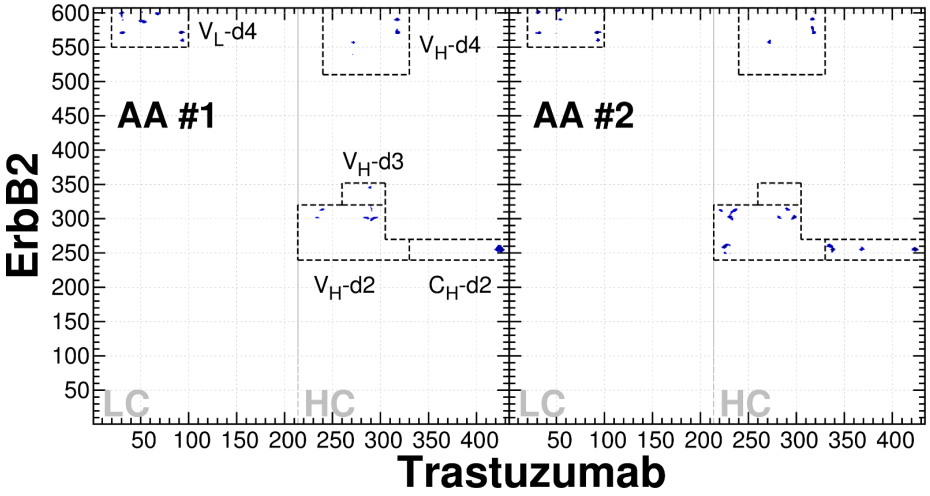


Figure 5.7: Distance matrices between ECD-ErbB2 and Trastuzumab-Fab for the atomistic clusters. Cluster 1 (*Left*) and cluster 2 (*Right*). The averaged interfaces are shown in dashed lines for V_L-d4, V_H-d4, V_H-d2, V_H-d3 and C_H-d2 respectively.

both clusters (See Figure 5.7). The main difference is that there are more residues from the subdomain II in cluster 2 than in cluster 1. About the interfaces V_H-d4 and V_L-d4 there are more overlapping in cluster 1 than in cluster 2, as can be seen in Figure 5.7. Another difference in the clusters mentioned above is the conformational disposition of ECD-ErbB2: cluster 2 shows contacts in its C-terminal coil, whereas cluster 1 shows a single contact in subdomain III. Therefore, taking into a count that the structures have been fitted on the Trastuzumab-Fab moiety, it can be seen that ECD-ErbB2 in cluster 1 is above the pseudo-plane described by the Fab while it is below that plane in cluster 2 (See Figure 5.5).

CG simulations

The first and second clusters of each CG simulation were used to compare with the AA simulation. For all CG methodologies the first two clusters describe more than 70 % of the trajectory as can be seen in Figure 5.14 and consequently are used for the subsequent analysis.

Figures 5.8 and 5.9 represent the distance matrices between ECD-ErbB2 and Trastuzumab-Fab for cluster 1 of each CG simulation (Figure 5.8 for $r_{list} = 1.2nm$ and Figure 5.9 for $r_{list} = 1.4nm$). The corresponding maps for cluster 2 are presented in Figures 5.15-5.16 in the Additional Information section. Each one is compared with cluster 1 and 2 from the AA simulation (Figure 5.7). In general, it can be observed that CG simulations enhance the contacts since CG tends to aggregate the proteins. Overall, CG simulations reproduce satisfactory well the AA pattern at the interfaces corresponding to the crystallographic contacts (V_L-

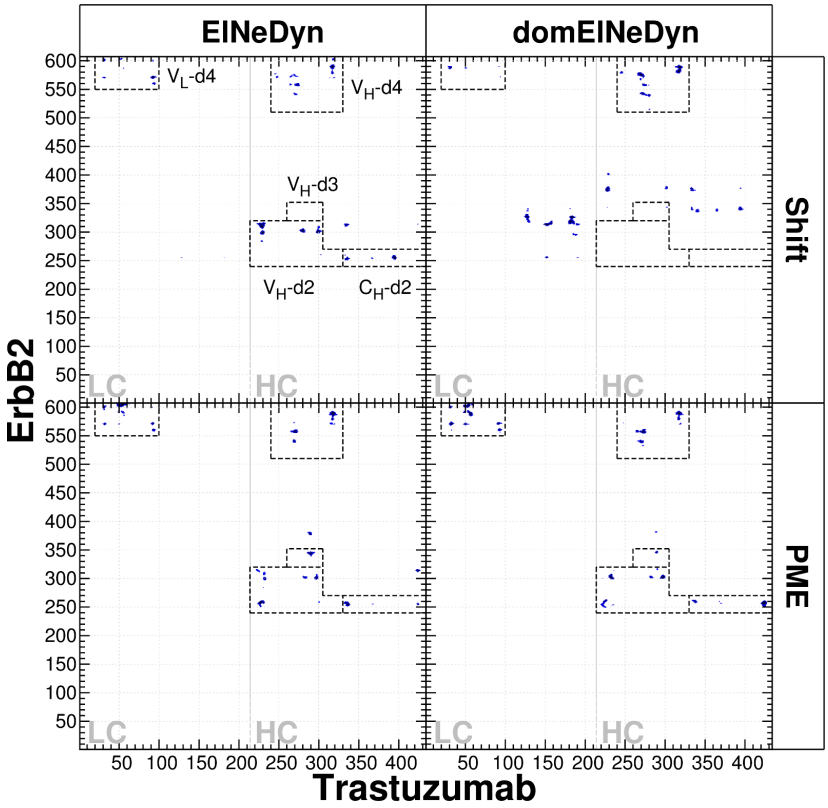


Figure 5.8: Distance matrices between ECD-ErbB2 and Trastuzumab-Fab for *cluster 1* corresponding to each Martini method with $r_{list} = 1.2nm$. The averaged interfaces are shown in dashed lines for V_L-d4, V_H-d4, V_H-d2, V_H-d3 and C_H-d2 respectively.

d4 and V_H-d4) except for those obtained with Shift and $r_{list} = 1.2nm$ (compare Figure 5.8 and 5.9). Looking at the new interface predicted by the AA simulation [16] using OPLS corresponding to the interactions between subdomains V_H-d2 and C_H-d2, there is not such a good correlation with the CG simulations. In Figures 5.8-5.9 it can be seen that P1.2, Pd1.2 and Sd1.4 described very well the interface between subdomains C_H-d2 and V_H-d2, but additional contacts appear on the map. It is also noticeable that cluster 1 of Pd1.4 shows a good correlation with cluster 2 of the AA simulation at both interfaces (See Figure 5.9). Although P1.2 and Pd1.2 have a good similarity with the AA pattern in the specific interface V_H-d2, Pd1.4 is still better when a detailed comparison is done.

The ErbB2-Trastuzumab contacts revealed by each AA cluster were merged and selected for the comparison with the CG simulations. The number of matches between the AA and CG contacts are expressed in % (where 100 % corresponds to hit all contacts) and shown in Figure 5.10. The plot represents the results

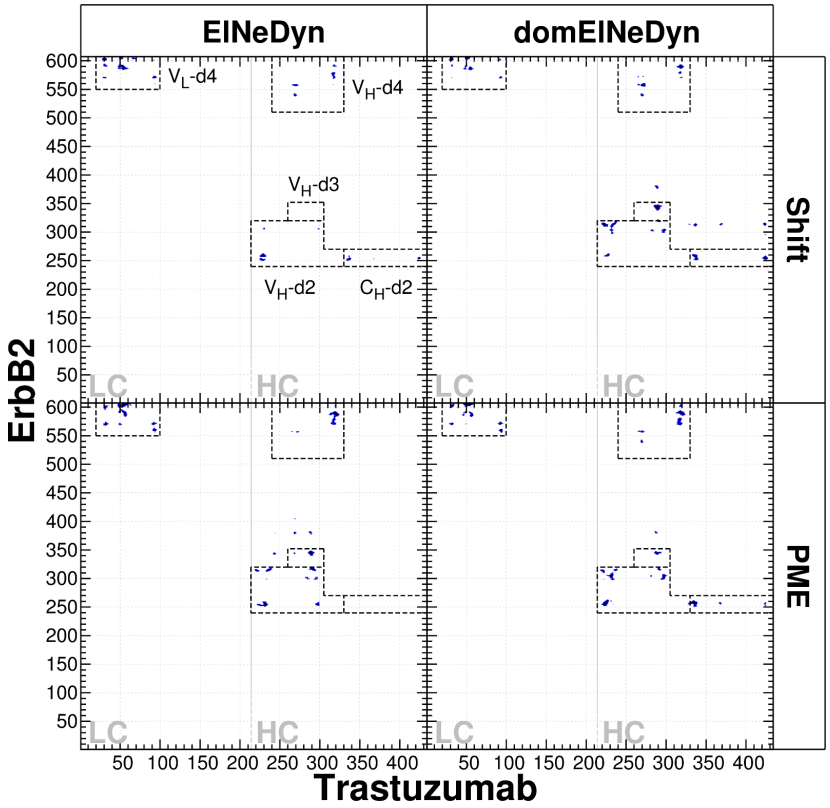


Figure 5.9: Distance matrices between ECD-ErbB2 and Trastuzumab-Fab for *cluster 1* corresponding to each Martini method with $r_{list} = 1.4nm$. The averaged interfaces are shown in dashed lines for V_L-d4, V_H-d4, V_H-d2, V_H-d3 and C_H-d2 respectively.

for the crystallographic contacts (V_L-d4 and V_H-d4) (black and blue bars) on one hand, and the additional contacts between C_H-d2 and V_H-d2 (red and cyan bars) on the other. The contacts between subdomains V_H-d3 (gray bar) involve the interactions between two pair of residues (see Table 5.10) that can be found in most CG methodologies, excepting the Sd1.2 and S1.2 simulations. Using the same r_{list} and EN protocol the best results for the crystallographic contacts is obtained when PME is applied. A better result can be reached using ElNeDyn than domElNeDyn when the same r_{list} and Shift EI method are used, whereas domElNeDyn achieves better results than ElNeDyn when the EI method is PME. The percentage of matches increases when r_{list} is reduced for the Shift models irrespective of the EN choice, whereas the same conclusion is not so clear for PME. On other hand, there is not a very clear trend in the V_H-d2 and C_H-d2 interfaces. However, for the same EI method and EN protocol the percentage of contacts increases with r_{list}. That tendency is more clear for Shift than for PME. In addition to this, it should be noticed that the CG models labeled as

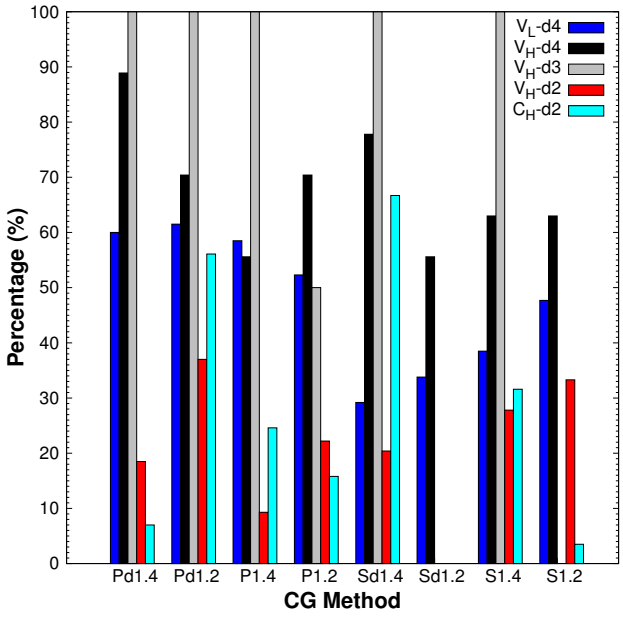


Figure 5.10: Histograms for the percentage of successful matches obtained with the Martini methods in both crystallographic (*blue and black*) and additional (*green, red and cyan*) contacts present in the atomistic clusters.

Pd1.2 and Sd1.4 reach a good correlation with the AA results in the V_H-d2 and C_H-d2 interfaces as shown in Figures 5.7-5.9. The Pd1.4 method shows a worse result than P1.2, Pd1.2 and Sd1.4 in the matched percentage (see Figure 5.10) in spite of having a very similar pattern regarding this interface (Figure 5.9). It is probably due to the displacement of one residue in the pair list as can be deduced from Figure 5.9.

5.4 Discussion

Based on the results reported in the previous section, a more detailed view of the interesting protein-protein interactions in the ECD-ErbB2/Trastuzumab-Fab complex is discussed in terms of the comparative performance of the CG methods with the AA simulation. The interactions showed in the crystallographic structure are analyzed first. In a second term, the additional interactions captured by the AA simulations are also discussed.

5.4.1 Crystallographic Interactions

The X-ray structure corresponds to an open complex which has a binding pocket involving the three loop regions of subdomain IV in ErbB2 with residues 557-561 (loop1), 570-573 (loop2) and 593-603 (loop3) and some loops in the variable subdomains of Trastuzumab, V_L and V_H [loop1-(V_H:57 and V_L:92-94); loop2-(V_H:103-105 and V_L:29-32,91-93); and loop3-(V_L:29-32, 66-68, 50-52)]. The first

Table 5.3: Percentage of native contacts in the crystallographic interface conserved in the CG simulations.

Label	Loop 1	Loop 2	Loop 3
Pd1.4	100.0	76.7	50.0
Pd1.2	90.9	50.0	40.0
P1.4	81.8	70.0	45.0
P1.2	63.6	46.7	60.0
Sd1.4	18.2	36.7	35.0
Sd1.2	0.0	30.0	25.0
S1.4	18.2	50.0	30.0
S1.2	81.8	50.0	35.0

and third loops have interactions primarily of electrostatic nature and the second loop makes mostly hydrophobic contacts, as was described by Cho and et al [29]. These crystallographic interactions have been kept along the AA simulation and by most of the CG simulations (See Figures 5.7-5.9). Nevertheless, in addition to those crystallographic contacts, the molecular dynamics on this system increase the interacting surface adding more residues at the interface between the V_L, V_H and IV subdomains. These findings were previously discussed [16]. The original binding interface reported in the X-ray structure is included in the native contacts considered in Figure 5.10 and listed in Tables 5.8-5.12. The calculated percentages of those native contacts involved in the binding interface that are matched in the CG simulations are shown in Table 5.3. It can be observed that Pd1.4 keeps the original binding interface better than the other CG methods. This is consistent with the fact that a larger r_{list} value (1.4 nm instead 1.2 nm) will take into account more interactions. In addition to this, the precision of the PME method to calculate the electrostatic interactions is well established [12, 13, 52, 53]. The worst result, which misses in fact most of the binding interface, is for the traditional Shift with r_{list} of 1.2 nm and an EN by domains.

5.4.2 Additional Protein-Protein Interaction

Franco et al. reported an additional interaction to that described in the X-ray structure between subdomain II and the subdomain C_H from Trastuzumab obtained by atomistic MD simulation. In reference [16] and in a recent work considering the full receptor in membrane simulated at a CG scale, this interaction was described [6]. The additional interface is mediated by subdomain II of the ECD-ErbB2 receptor, primarily by the dimerization arm and the subdomain C_H of the Trastuzumab-Fab. This happens due to a hinge movement intrinsic to the EGFR receptor in its ectodomain as has been reported previously [51]. As can be seen in Tables 5.8-5.12 of the Additional Information section the additional

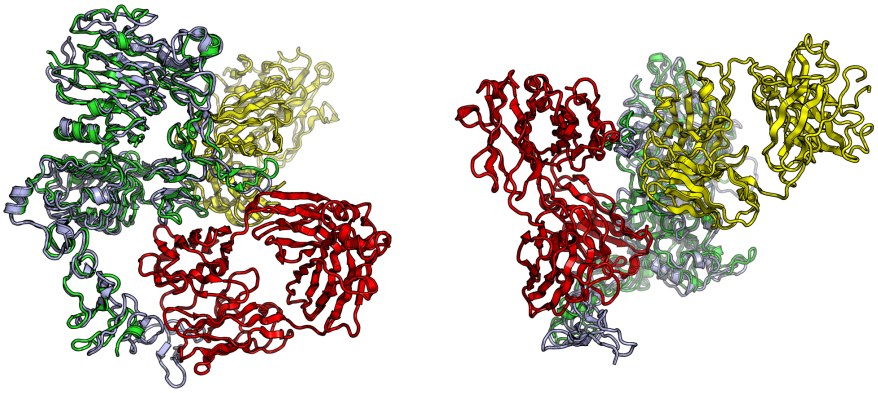


Figure 5.11: The structures corresponding to cluster 1 that have been fitted to the x-ray structure of the complex ECD-ErbB2/pertuzumab identified as PDB code 1S78. *Blue and red* are ECD-ErbB2/Trastuzumab-Fab from **cluster 1** while *green and yellow* are ECD-ErbB2/Pertuzumab-Fab

interface of subdomain II and the constant domain involves residues 253-261 of ErbB2 (where residues 259-261 correspond to a sheet) and the residues 121, 123-125, 153-154, 206-213 of Trastuzumab (where residues 208-211 form an helix and 151-152, 206-207 form sheets). This description reveals an interface with a mixing of hydrophobic and electrostatic contacts.

In view of these results, we can speculate about the potent anti-ErbB2 characteristic of Trastuzumab [15, 54, 55] because the antibody constant domain is able to interact with the dimerization arm. It has been reported that an antibody can increase its therapeutic potential if it is able to also bind by the constant domain [56] in addition to the ability of the variable Fab domain to recognize the surface of the antigen (epitope) [57]. The additional interaction suggests new questions regarding the accessibility of the dimerization arm and their biological implications for other antibodies. It has been reported that the usage of a combinatory therapy with two antibodies, pertuzumab and trastuzumab results in a better clinical outcome [58–63]. The Pertuzumab target is the ErbB2 dimerization domain. In fact, a crystallographic structure of the complex ECD-ErbB2/Pertuzumab has been reported and identified as 1S78 in the PDB [36]. Therefore, we were wondering whether the additional ErbB2-Trastuzumab interface observed in the AA and CG simulations is compatible with the ErbB2-Pertuzumab interaction or with the dimerization activity. Looking to our molecular dynamics structures, the presence of both interacting antibodies with the ErbB2 ectodomain is well-matched. The binding region corresponding to the pertuzumab epitope is placed in the opposite side to the ErbB2-Trastuzumab V_H-d2 and C_H-d2 interface. Fig-

ure 5.11 shows the atomistic cluster 1 structure fitted to the crystallographic ECD-ErbB2-pertuzumab complex (PDB: 1S78). The RMSD of the fitting performed onto the ECD-ErbB2 backbone atoms was only 0.25 nm, indicating an acceptable match.

5.5 Conclusions From This Work

We have reported a comparative study between the Martini CG and the atomistic OPLS-AA models applied to the MD simulation of the ECD-ErbB2/Trastuzumab-Fab complex. In particular, the following CG settings have been evaluated: the effect of the cutoff radius to find neighbors (r_{list}), the influence of the method to calculate the electrostatic interactions (PME or Shift) and the usage the Elastic Networks (an elastic network for ErbB2 protein and an elastic network by each chain of Trastuzumab-Fab) either by domains or just for the w hole protein (domElNeDyn or ElNeDyn).

In general the CG simulations are able to satisfactorily reproduce the crystallographic determined interaction contacts between the ECD-ErbB2 and the Trastuzumab-Fab proteins as well as the additional interface involving the ErbB2 subdomain II already described in the atomistic simulations. The PME treatment of electrostatic interactions seems to behave as good as or even better than the Shift methodology in view of the different results reported in this chapter. The ElNeDyn protocol for the treatment of elastic networks shows more flexibility when the networks are built by domains instead of considering the whole protein. It allows exploring a wider region of the conformational space at the expense of some mismatch with the atomistic models.

It also can be concluded that the larger neighbor lists cut-off radius of 1.4 nm gives better results when compared with the atomistic system. The combination of PME, r_{list} of 1.4 nm and an EN by domains gives an overall satisfactory correlation with the atomistic simulation structures, in particular, the protein-protein interface has been captured adequately by the CG model. This result can have important consequences regarding the consideration of using GPU technologies. Currently, the Gromacs package (and most of other packages) only implements PME for electrostatic interaction calculations in the GPU versions. Keeping in mind, that originally Martini CG models were developed using the Shift method in the electrostatic calculations, it is important to check that PME is suitable for CG simulations on GPUs.

5.6 *Additional Information*

Secondary Structure Analysis

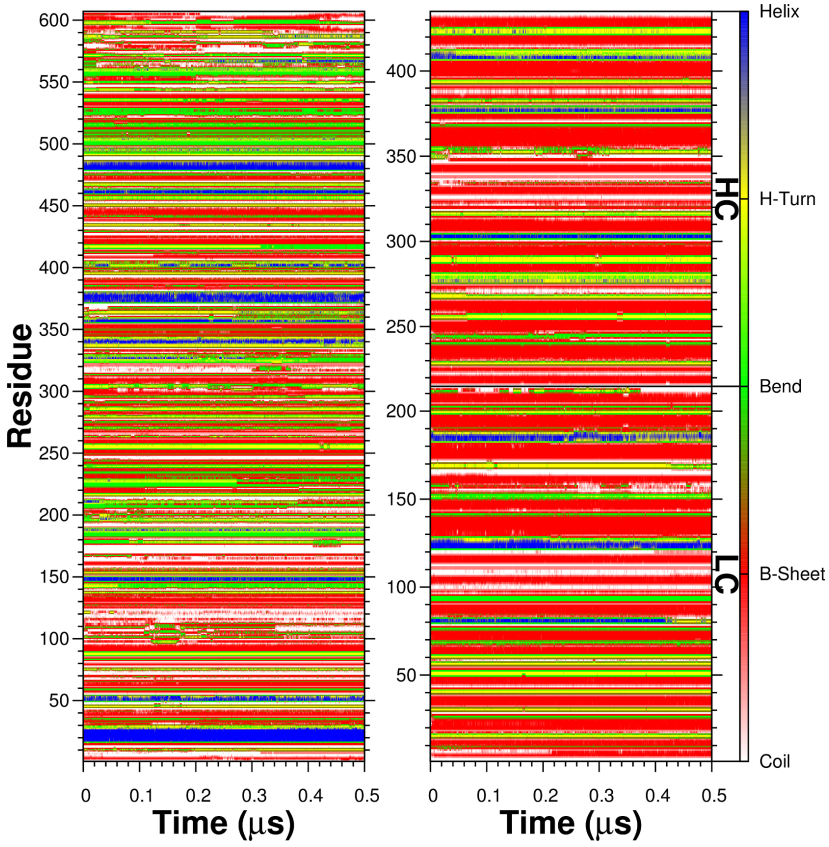


Figure 5.12: Analysis DSSP for AA simulation. *Left*, ecto-ErbB2 and *Right*, Trastuzumab-Fab

RMSF for CG Trajectories

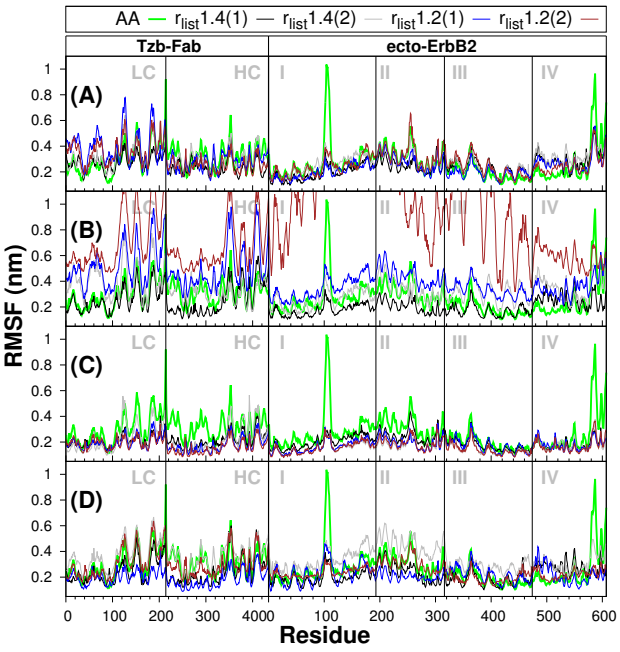


Figure 5.13: RMSF per residue for the BB particles of the Martini and Cα of the atomistic simulations (green line). Minimized atomistic structure converted to CG was used as the reference structure. (A) Shift-ElNeDyn methods, (B) Shift-domElNeDyn methods, (C) PME-ElNeDyn methods, and (D) PME-domElNeDyn

Cluster Size

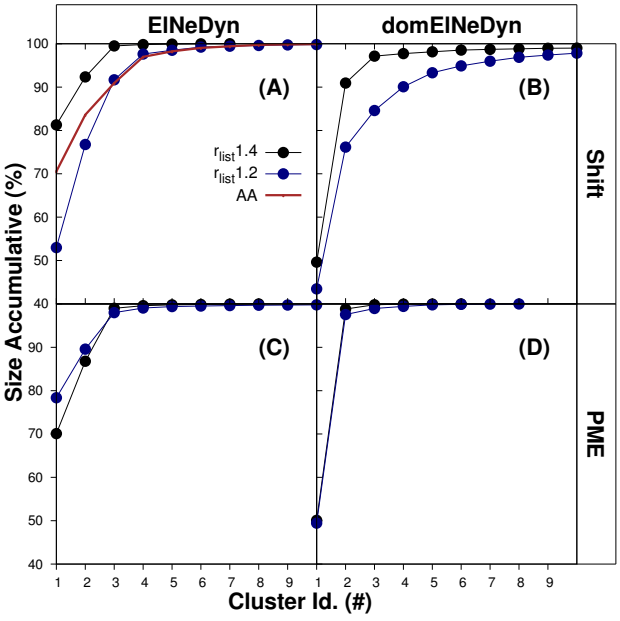


Figure 5.14: Cluster Size Accumulative distribution by CG simulations methodologies. (A) Shift-ElNeDyn and AA simulation, (B) Shift-domElNeDyn, (C) PME-ElNeDyn and (D) PME-domElNeDyn.

Distance Matrices

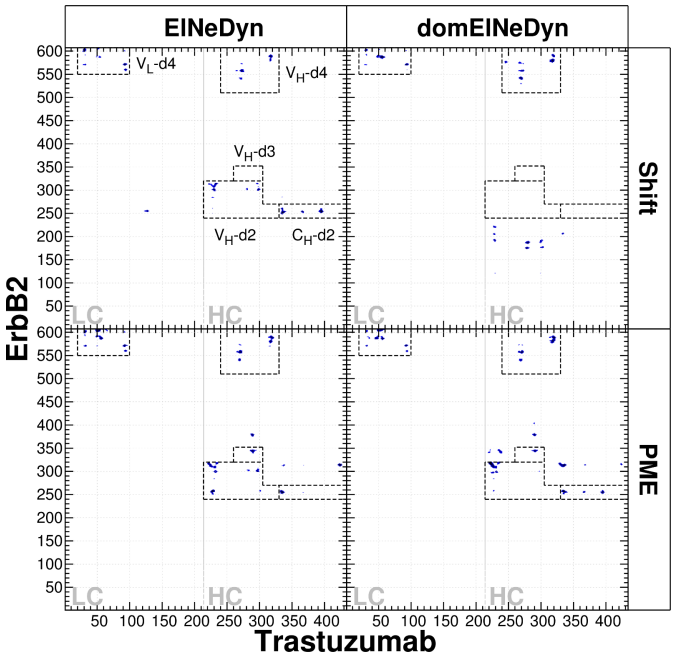


Figure 5.15: Distance matrices between ECD-ErbB2 and Trastuzumab-Fab for *cluster 2* corresponding to each Martini method with $r_{list} = 1.2nm$. The averaged interfaces are shown in dashed lines for V_L-d4, V_H-d4, V_H-d2, V_H-d3 and C_H-d2 respectively.

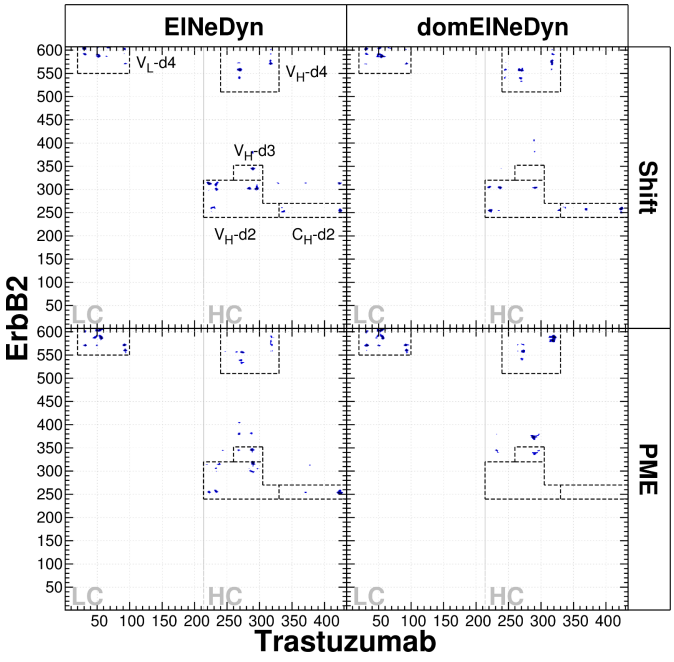


Figure 5.16: Distance matrices between ECD-ErbB2 and Trastuzumab-Fab for *cluster 2* corresponding to each Martini method with $r_{list} = 1.4nm$. The averaged interfaces are shown in dashed lines for V_L-d4, V_H-d4, V_H-d2, V_H-d3 and C_H-d2 respectively.

RMSD Mean

Table 5.4: Average RMSD on the last 200 ns at atomistic simulation time window

Label	Replica	Mean	SD
AA	n.a	0.90	0.05
P1.2	1	0.91	0.05
	2	1.03	0.04
P1.4	1	0.83	0.04
	2	0.98	0.03
Pd1.2	1	1.04	0.04
	2	1.10	0.08
Pd1.4	1	0.86	0.02
	2	1.09	0.05
S1.2	1	1.01	0.05
	2	1.01	0.04
S1.4	1	0.85	0.06
	2	1.05	0.05
Sd1.2	1	1.58	0.06
	2	1.53	0.28
Sd1.4	1	1.17	0.04
	2	1.22	0.06

Table 5.5: Average RMSD on the last 16 μs at μ-second timescale

Label	Replica	Mean	SD
P1.2	1	0.96	0.07
	2	1.11	0.04
P1.4	1	0.84	0.04
	2	0.94	0.06
Pd1.2	1	1.16	0.05
	2	1.29	0.03
Pd1.4	1	0.95	0.05
	2	1.12	0.06
S1.2	1	0.92	0.04
	2	1.23	0.07
S1.4	1	0.92	0.05
	2	0.91	0.07
Sd1.2	1	2.09	0.04
	2	2.55	0.06
Sd1.4	1	1.28	0.05
	2	1.14	0.02

RMSIP Analysis

Table 5.6: RMSIP for the different CG methods at atomistic simulation time window

Label	Mean	SD
P1.2	0.68	< 0.0
P1.4	0.70	< 0.0
Pd1.2	0.67	0.01
Pd1.4	0.69	0.01
S1.2	0.67	< 0.0
S1.4	0.70	< 0.0
Sd1.2	0.68	0.01
Sd1.4	0.69	0.01

Table 5.7: RMSIP for the different CG methods at μ-second timescale

Label	Mean	SD
P1.2	0.66	< 0.0
P1.4	0.68	0.02
Pd1.2	0.66	< 0.0
Pd1.4	0.66	0.02
S1.2	0.67	0.03
S1.4	0.67	0.01
Sd1.2	0.50	0.02
Sd1.4	0.68	0.01

Pair Lists Involved in the Interaction Patterns

Table 5.8: Residues Involved in the Interface V_L-d4.

Pair List									
29 VAL	571 PRO	31 THR	600 ALA	50 SER	602 GLN	68 GLY	599 GLY	93 THR	571 PRO
29 VAL	600 ALA	31 THR	601 CYS	50 SER	603 PRO	68 GLY	600 ALA	93 THR	572 PRO
29 VAL	570 ASP	31 THR	570 ASP	50 SER	604 CYS	91 HIS	571 PRO	93 THR	573 PHE
29 VAL	572 PRO	31 THR	602 GLN	52 SER	587 SER	91 HIS	572 PRO	93 THR	559 ALA
29 VAL	601 CYS	31 THR	603 PRO	52 SER	588 TYR	91 HIS	570 ASP	93 THR	561 GLN
30 ASN	571 PRO	32 ALA	571 PRO	52 SER	603 PRO	91 HIS	573 PHE	93 THR	562 CYS
30 ASN	599 GLY	32 ALA	572 PRO	53 PHE	587 SER	92 TYR	560 ASP	93 THR	570 ASP
30 ASN	600 ALA	32 ALA	570 ASP	53 PHE	588 TYR	92 TYR	570 ASP	94 THR	560 ASP
30 ASN	601 CYS	32 ALA	601 CYS	53 PHE	589 MET	92 TYR	571 PRO	94 THR	561 GLN
30 ASN	570 ASP	50 SER	588 TYR	54 LEU	587 SER	92 TYR	572 PRO	94 THR	572 PRO
30 ASN	602 GLN	50 SER	589 MET	54 LEU	588 TYR	92 TYR	573 PHE	94 THR	558 GLU
30 ASN	603 PRO	50 SER	590 PRO	66 ARG	599 GLY	92 TYR	559 ALA	94 THR	559 ALA
31 THR	571 PRO	50 SER	592 TRP	67 SER	599 GLY	93 THR	560 ASP	94 THR	562 CYS

Table 5.9: Residues Involved in the Interface V_H-d4.

Pair List									
57 TYR	557 PRO	102 ASP	580 SER	103 GLY	573 PHE	103 GLY	590 PRO	104 PHE	591 ILE
57 TYR	558 GLU	102 ASP	590 PRO	103 GLY	577 ARG	103 GLY	591 ILE	105 TYR	571 PRO
101 GLY	591 ILE	102 ASP	591 ILE	103 GLY	578 ARG	103 GLY	592 TRP	105 TYR	572 PRO
102 ASP	577 ARG	103 GLY	570 ASP	103 GLY	579 PRO	103 GLY	593 LYS		
102 ASP	578 ARG	103 GLY	571 PRO	103 GLY	580 SER	104 PHE	571 PRO		
102 ASP	579 PRO	103 GLY	572 PRO	103 GLY	589 MET	104 PHE	590 PRO		

Table 5.10: Residues Involved in the Interface V_H-d3.

Pair List	
75 SER	346 LYS
75 SER	347 LYS

Table 5.11: Residues Involved in the Interface V_H-d2.

Pair List									
7 SER	313 SER	13 GLN	262 ASN	67 ARG	303 GLU	76 LYS	314 LYS	83 MET	303 GLU
9 GLY	259 SER	15 GLY	303 GLU	69 THR	302 ALA	76 LYS	315 PRO	84 ASN	301 THR
10 GLY	259 SER	16 GLY	302 ALA	70 ILE	302 ALA	76 LYS	316 CYS	84 ASN	302 ALA
11 LEU	258 GLU	16 GLY	303 GLU	71 SER	301 THR	78 THR	313 SER	84 ASN	303 GLU
11 LEU	259 SER	17 SER	302 ALA	71 SER	302 ALA	78 THR	314 LYS	84 ASN	304 ASP
11 LEU	260 MET	17 SER	303 GLU	73 ASP	314 LYS	80 TYR	302 ALA	85 SER	303 GLU
11 LEU	261 PRO	18 LEU	301 THR	76 LYS	298 GLN	81 LEU	302 ALA	119 ARG	261 PRO
12 VAL	259 SER	18 LEU	302 ALA	76 LYS	299 GLU	82 GLN	301 THR	120 SER	261 PRO
12 VAL	260 MET	19 ARG	310 GLU	76 LYS	311 LYS	82 GLN	302 ALA	120 SER	262 ASN
12 VAL	261 PRO	19 ARG	311 LYS	76 LYS	312 CYS	83 MET	301 THR	120 SER	263 PRO
13 GLN	261 PRO	20 LEU	302 ALA	76 LYS	313 SER	83 MET	302 ALA		

Table 5.12: Residues Involved in the Interface C_H-d2.

Pair List									
121 ALA	261 PRO	153 PHE	255 ASP	208 LYS	258 GLU	210 SER	255 ASP	211 ASN	258 GLU
123 THR	253 ASN	153 PHE	256 THR	208 LYS	259 SER	210 SER	256 THR	211 ASN	259 SER
123 THR	254 THR	154 PRO	256 THR	209 PRO	253 ASN	210 SER	257 PHE	211 ASN	260 MET
123 THR	255 ASP	206 ASN	256 THR	209 PRO	254 THR	210 SER	258 GLU	212 THR	254 THR
123 THR	256 THR	207 HIS	255 ASP	209 PRO	255 ASP	210 SER	259 SER	212 THR	255 ASP
123 THR	258 GLU	207 HIS	256 THR	209 PRO	256 THR	210 SER	260 MET	212 THR	256 THR
123 THR	260 MET	207 HIS	257 PHE	209 PRO	257 PHE	210 SER	261 PRO	212 THR	257 PHE
123 THR	261 PRO	208 LYS	253 ASN	209 PRO	258 GLU	211 ASN	253 ASN	213 LYS	255 ASP
124 LYS	255 ASP	208 LYS	254 THR	209 PRO	259 SER	211 ASN	254 THR	213 LYS	256 THR
124 LYS	256 THR	208 LYS	255 ASP	209 PRO	260 MET	211 ASN	255 ASP		
125 GLY	255 ASP	208 LYS	256 THR	210 SER	253 ASN	211 ASN	256 THR		
125 GLY	256 THR	208 LYS	257 PHE	210 SER	254 THR	211 ASN	257 PHE		

References

[1] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. *The Journal of Physical Chemistry B*, 111:7812–7824, 2007.

[2] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. *Journal of Chemical Theory and Computation*, 4:819–834, 2008.

[3] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A. Ceruso. *Journal of Chemical Theory and Computation*, 5:2531–2543, 2009.

[4] Marc Baaden and Siewert J. Marrink. *Current Opinion in Structural Biology*, 23:878–886, 2013.

[5] Austin C. Stark, Casey T. Andrews, and Adrian H. Elcock. *Journal of Chemical Theory and Computation*, 9:4176–4185, 2013.

[6] J. F. Franco-Gonzalez, J. Ramos, V. L. Cruz, and J. Martinez-Salazar. *J Comput Aided Mol Des*, pages 1–15, 2014.

[7] Iwona Siuda and Lea Thgersen. *Journal of Molecular Modeling*, 19:4931–4945, 2013.

[8] R. L. Marsden, L. J. McGuffin, and D. T. Jones. *Protein Sci*, 11:2814–24, 2002.

[9] Djurre H. de Jong, Gurpreet Singh, W. F. Drew Bennett, Clement Arnarez, Tsjerk A. Wassenaar, Lars V. Schfer, Xavier Periole, D. Peter Tieleman, and Siewert J. Marrink. *Journal of Chemical Theory and Computation*, 9:687–697, 2012.

[10] Modesto Orozco. *Chemical Society Reviews*, 43:5051–5066, 2014.

[11] Fabio Sterpone, Simone Melchionna, Pierre Tuffery, Samuela Pasquali, Normand Mousseau, Tristan Cragnolini, Yassmine Chebaro, Jean-Francois St-Pierre, Maria Kalimeri, Alessandro Barducci, Yoann Laurin, Alex Tek, Marc Baaden, Phuong Hoang Nguyen, and Philippe Derreumaux. *Chemical Society Reviews*, 43:4871–4893, 2014.

[12] T. Darden, D. York, and L. Pedersen. *J. Chem. Phys.*, 98:10089–10092, 1993.

[13] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. *J. Chem. Phys.*, 103:8577–8593, 1995.

[14] Xavier Periole and Siewert-Jan Marrink. *The Martini Coarse-Grained Force Field*, volume 924 of *Methods in Molecular Biology*, pages 533–565. Humana Press, 2013.

[15] Gloria Fuentes, Maurizio Scaltriti, Jose Baselga, and Chandra Verma. *Breast Cancer Research*, 13:R54, 2011.

[16] J. F. Franco-Gonzalez, J. Ramos, V. L. Cruz, and J. Martinez-Salazar. *J. Mol. Model.*, 19:1227–1236, 2013.

[17] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F. van Gunsteren, and Alan E. Mark. *Angewandte Chemie International Edition*, 38:236–240, 1999.

[18] A. Amadei, A. B. Linssen, B. L. de Groot, D. M. van Aalten, and H. J. Berendsen. *J Biomol Struct Dyn*, 13:615–25, 1996.

[19] R. Patel and H. Y. Leung. *Curr Pharm Des*, 18:2672–9, 2012.

[20] J. Baselga and S. M. Swain. *Nat Rev Cancer*, 9:463–75, 2009.

[21] D. J. Slamon, G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, and W. L. McGuire. *Science*, 235:177–82, 1987.

[22] J. Baselga, D. Tripathy, J. Mendelsohn, S. Baughman, C. C. Benz, L. Dantis, N. T. Sklarin, A. D. Seidman, C. A. Hudis, Moore J, Rosen Pp, Twaddell T, Henderson Ic, and Norton L. *J. Clin. Oncol.*, 14:737–737, 1996.

[23] P. Carter, L. Presta, C. M. Gorman, J. B. Ridgway, D. Henner, W. L. Wong, A. M. Rowland, C. Kotts, M. E. Carver, and H. M. Shepard. *Proc Natl Acad Sci U S A*, 89:4285–9, 1992.

[24] Charles L. Vogel, Melody A. Cobleigh, Debu Tripathy, John C. Gutheil, Lyndsay N. Harris, Louis Fehrenbacher, Dennis J. Slamon, Maureen Murphy, William F. Novotny, Michael Burchmore, Steven Shak, Stanford J.

Stewart, and Michael Press. *Journal of Clinical Oncology*, 20:719–726, 2002. 10.1200/JCO.20.3.719.

[25] Rita Nahta and Francisco J. Esteva. *Cancer Letters*, 232:123–138, 2006.

[26] R. Nahta and F. J. Esteva. *Oncogene*, 26:3637–3643, 2007.

[27] R. Nahta. *Current Medicinal Chemistry*, 19:1065–10755, 2012.

[28] Miguel A. Molina, Jordi Codony-Servat, Joan Albanell, Federico Rojo, Joaqun Arribas, and Jose Baselga. *Cancer Research*, 61:4744–4749, 2001.

[29] Hyun-Soo Cho, Karen Mason, Kasra X. Ramyar, Ann Marie Stanley, Sandra B. Gabelli, Dan W. Denney, and Daniel J. Leahy. *Nature*, 421:756–760, 2003.

[30] Thomas P. J. Garrett, Neil M. McKern, Meizhen Lou, Thomas C. Elleman, Timothy E. Adams, George O. Lovrecz, Michael Kofler, Robert N. Jorissen, Edouard C. Nice, Antony W. Burgess, and Colin W. Ward. *Molecular Cell*, 11:495–505, 2003.

[31] Thomas P. J. Garrett, Neil M. McKern, Meizhen Lou, Thomas C. Elleman, Timothy E. Adams, George O. Lovrecz, Hong-Jian Zhu, Francesca Walker, Morry J. Frenkel, Peter A. Hoyne, Robert N. Jorissen, Edouard C. Nice, Antony W. Burgess, and Colin W. Ward. *Cell*, 110:763–773, 2002.

[32] Kathryn M. Ferguson, Mitchell B. Berger, Jeannine M. Mendrola, Hyun-Soo Cho, Daniel J. Leahy, and Mark A. Lemmon. *Molecular Cell*, 11:507–517, 2003.

[33] Hyun-Soo Cho and Daniel J. Leahy. *Science*, 297:1330–1333, 2002.

[34] Samuel Bouyain, Patti A. Longo, Shiqing Li, Kathryn M. Ferguson, and Daniel J. Leahy. *Proceedings of the National Academy of Sciences of the United States of America*, 102:15024–15029, 2005.

[35] Chafen Lu, Li-Zhi Mi, Michael J. Grey, Jieqing Zhu, Elizabeth Graef, Shigeyuki Yokoyama, and Timothy A. Springer. *Molecular and Cellular Biology*, 30:5432–5443, 2010.

[36] Matthew C. Franklin, Kendall D. Carey, Felix F. Vajdos, Daniel J. Leahy, Abraham M. de Vos, and Mark X. Sliwkowski. *Cancer Cell*, 5:317–328, 2004.

[37] Juan F. Franco-Gonzalez, Victor Cruz, Javier Ramos, and Javier Martinez-Salazar. *Biophysical Journal*, 106:666a–667a, 2014.

[38] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. *Bioinformatics*, 29:845–54, 2013.

[39] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. *J. Chem. Phys.*, 81:3684–3690, 1984.

[40] W. L. Jorgensen and J. Tirado-Rives. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[41] H. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. *Intermolecular forces*, pages 331–342. Reidel, Dordrecht, 1981.

[42] B. Hess, H. Bekker, H. J. C. Berendsen, and Jgem Fraaije. *J. Comput. Chem.*, 18:1463–1472, 1997.

[43] Siewert J. Marrink, Alex H. de Vries, and Alan E. Mark. *The Journal of Physical Chemistry B*, 108:750–760, 2004.

[44] Antony W. Burgess, Hyun-soo Cho, Charles Eigenbrot, Kathryn M. Ferguson, Thomas P. J. Garrett, Daniel J. Leahy, Mark A. Lemmon, Mark X. Sliwkowski, Colin W. Ward, and Shigeyuki Yokoyama. *Molecular Cell*, 12:541–552, 2003.

[45] [cited 09/2014]. Available from: http://www.reading.ac.uk/bioinf/domfold, 2014.

[46] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. *Proteins Struct. Funct. Bioinforma.*, 17:412–425, 1993.

[47] B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. Caves. *Bioinformatics*, 22:2695–6, 2006.

[48] R. Elber and M. Karplus. *Science*, 235:318–321, 1987.

[49] Alemayehu A. Gorfe, Riccardo Baron, and J. Andrew McCammon. *Biophysical Journal*, 95:3269–3277, 2008.

[50] Peter L. Freddolino, Feng Liu, Martin Gruebele, and Klaus Schulten. *Biophysical Journal*, 94:L75–L77, 2008.

[51] Juan Felipe Franco-Gonzalez, Javier Ramos, Victor L. Cruz, and Javier Martnez-Salazar. *Journal of Molecular Modeling*, 19:931–941, 2013.

[52] Celeste Sagui and Thomas A. Darden. *Annual Review of Biophysics and Biomolecular Structure*, 28:155–179, 1999.

[53] Stewart A. Adcock and J. Andrew McCammon. *Chemical Reviews*, 106:1589–1615, 2006.

[54] S. L. Morrison, M. J. Johnson, L. A. Herzenberg, and V. T. Oi. *Proceedings of the National Academy of Sciences*, 81:6851–6855, 1984.

[55] Genevieve Hansen. Antibody constant domain regions and uses thereof, 2014.

[56] Felix F. Vajdos, Camellia W. Adams, Timothy N. Breece, Leonard G. Presta, Abraham M. de Vos, and Sachdev S. Sidhu. *Journal of Molecular Biology*, 320:415–428, 2002.

[57] Ken Muphy, Paul Travers, and Mark Walpor. *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science, 2001.

[58] Yoichi Nagata, Keng-Hsueh Lan, Xiaoyan Zhou, Ming Tan, Francisco J. Esteva, Aysegul A. Sahin, Kristine S. Klos, Ping Li, Brett P. Monia, Nina T. Nguyen, Gabriel N. Hortobagyi, Mien-Chie Hung, and Dihua Yu. *Cancer Cell*, 6:117–127, 2004.

[59] Richard M. Neve, Koei Chin, Jane Fridlyand, Jennifer Yeh, Frederick L. Baehner, Tea Fevr, Laura Clark, Nora Bayani, Jean-Philippe Coppe, Frances Tong, Terry Speed, Paul T. Spellman, Sandy DeVries, Anna Lapuk, Nick J. Wang, Wen-Lin Kuo, Jackie L. Stilwell, Daniel Pinkel, Donna G. Albertson, Frederic M. Waldman, Frank McCormick, Robert B. Dickson, Michael D. Johnson, Marc Lippman, Stephen Ethier, Adi Gazdar, and Joe W. Gray. *Cancer Cell*, 10:515–527, 2006.

[60] Rita Nahta, Linda X. H. Yuan, Bing Zhang, Ryuji Kobayashi, and Francisco J. Esteva. *Cancer Research*, 65:11118–11128, 2005.

[61] Jeffrey Ross, Elzbieta Slodkowska, W. F. Symmans, Lajos Pusztai, Peter Ravdin, and Gabriel Hortobagyi. *The oncologist*, 14:320–368, 2009.

[62] Luca Gianni, Tadeusz Pienkowski, Young-Hyuck Im, Laslo Roman, Ling-Ming Tseng, Mei-Ching Liu, Ana Lluch, Elbieta Staroslawska, Juan de la Haba-Rodriguez, Seock-Ah Im, Jose Luiz Pedrini, Brigitte Poirier, Paolo Morandi, Vladimir Semiglazov, Vichien Srimuninnimit, Giulia Bianchi, Tania Szado, Jayantha Ratnayake, Graham Ross, and Pinuccia Valagussa. *The Lancet Oncology*, 13:25–32, 2012.

[63] Javier Cortes, Sung-Bae Kim, Seock-Ah Im, Roberto Hegg, Jos Baselga, Javier Corts, Young-Hyuck Im, Laslo Roman, Jos Pedrini, Tadeusz Pienkowski, Adam Knott, Emma Clark, Mark Benyunes, Graham Ross, and Sandra Swain. *The New England journal of medicine*, 366:109–119, 2012.

Chapter 6

Full–Length ErbB2 Receptor Description

This chapter is based on the published article *"Exploring the dynamics and interaction of a full ErbB2 receptor and Trastuzumab-Fab antibody in a lipid bilayer model using Martini coarse-grained force field"* by JF Franco-Gonzalez, J Ramos, VL Cruz and J Martínez-Salazar, *J. Comput. Aided Mol. Des.*, 28(11):1093-1107, 2014

Abstract

Coarse Grained modeling has been applied to study the influence of the Trastuzumab monoclonal antibody on the structure and dynamics of the full ErbB2 receptor dimer, including the lipid bilayer. We will show that the Martini model performs satisfactorily well, giving results well-matched with those obtained by atomistic models as well as with the experimental information existing on homolog receptors. The Trastuzumab Fab hinders the interaction of the receptors with the lipid bilayer and generate the disruption of the antiparallel arrangement of the juxtamembrane segments in the dimer case.

6.1 Introduction

Most of the structural studies, either computational or experimental, performed on the EGFR family members have been carried out separately on ECD[1–6], transmembrane[7–17] and TKD[18–29]. The inherent difficulties associated to both the presence of the membrane and the size of the full systems are the main reason of such fragmented structural analysis. Several authors have used Fluorescence Resonance Energy Transfer (FRET) techniques to study the disposition of the full receptors in the membrane[30–32]. Some donor labels are placed in the protein structure and some acceptor labels in the lipid polar heads, allowing the measurement of the distance between the labeled residues and the lipid bilayer surface. Some set of FRET measurements using a labeled Trastuzumab Fab along with homology modeling techniques allowed Bagossi et al. to propose a structural model for the full ErbB2/Trastuzumab-Fab dimer inserted in the lipid bilayer[33].

There are only two molecular simulations studies dealing with an EGFR receptor inserted in the membrane. Kastner et al. studied the solvated EGFR tetramer in a lipid bilayer[30]. The study considers the extracellular and transmembrane domains of the receptors summing up to more than 8x105 atoms. The multinanosecond simulations (∼75ns) were performed using 4096 processors of the BlueGene supercomputer. Another related system studied by atomistic molecular dynamics simulation is the near half million atom complex simulated over tens of μs by Shaw's group on the Anton supercomputer, a "hardware" specifically designed to accommodate molecular dynamics algorithms. This system consisted of a complete EGFR dimer including the lipid bilayer[34].

Both studies used massive supercomputer resources which are hardly accessible to the scientific community. The coarse grained (CG) models offer an alternative way to study both larger spatial and time scales of huge biomolecular systems, beyond what is possible with all-atom models (AA), using moderate computational resources[35]. In particular, the Martini force field has been established as a suitable model for the coarse grained description of biological systems[36, 37]. Four atoms per coarse grained particle are mapped in Martini, which offers parameters for several types of molecules including, essentially, lipids and proteins. This model has been successfully used to investigate the structure and dynamics of systems containing lipids and peptides such as transmembrane proteins or antimicrobial peptides[38–40].

Here, we report on the molecular dynamics simulations of the structure and dynamics of the ErbB2 receptor and its interaction with the Trastuzumab monoclonal antibody Fab(mAb) from the experimental and computational points of view [41–43]. The ErbB2 receptor is studied in monomer and dimer forms, for which existence experimental evidence has been reported [44, 45]. The dimer is associated to receptor overexpressed in cancer cells[45]. Thus, in this manuscript

we analyze the capability of the Martini model to perform coarse grained simulation of interaction of a full ErbB2 receptor and Fab-Trastuzumab comparing the results with those obtained by atomistic modeling and also with other experimental data available on similar systems. It should be also pointed out that the simulation of these ErbB2 full receptor systems in its monomeric and dimeric forms, including the lipid bilayer and an antibody Fab has not been performed before.

6.2 Computational Methods

Protein – Bilayer Setup

An atomistic model of the full ErbB2 receptor and the Trastuzumab Fab were taken from the homology structure published by Bagossi [33]. Tables 6.3 – 6.5 in the section 6.9 provides the amino acid sequences of the different receptor and the antibody domains used in the present work. The Martini Force Field and its extension to proteins (version 2.2) were used to build the CG topology[36, 37]. Elastic Networks (EN) for the CG model of the protein in solution was employed using the ElNeDyn method[46].

The application of EN to the whole receptor might introduce undesirable rigidity to the system. The subdomain selection for the EN model is based on the receptor structural characteristics and our previous results on the atomistic ECD. For example, the extracellular subdomains I and III consists of a series of β-sheets which confer a relative rigidity to them. Subdomains II and IV are cystein-rich regions which may be assumed to be also relatively stable in water solution. These assumptions are confirmed by the different ErbBs structures experimentally elucidated either in the staggered or opened conformations simulations[30, 34, 41, 42, 47]. Therefore, the major sources of flexibility may be associated, in principle, with the interdomain connections. However, a validation of the EN setting procedure was performed by comparison between coarse-grained and atomistic simulations.

Elastic networks between backbone particles were applied separately at each subdomain for the ErbB2 complex and for the Trastuzumab Fab. The parameters controlling the elastic networks were a cutoff radius $R_c = 0.8nm$ and a spring constant $K_{spring} = 500kJmol^{-1}nm^{-2}$. We followed the methodology suggested by Periole et al.[46] regarding the selection of parameters for the EN. In the Figure 6.16 of section 6.9, the root mean squared fluctuation (RMSF) values corresponding to the ECD are represented for the different combinations of R_c and K_{spring} values compared with the atomistic data. As can be observed the best fit with the atomistic model is obtained with the combination R_c and K_{spring} suggested above.

We have selected the Dipalmitoylphosphatidylcholine (DPPC) lipid as a model

Table 6.1: Number of replicas and simulation time for each system

System	# Replicas	MD production[a](μs)
Monomer	3	4
Monomer+Tzb	2	4
Dimer	3	4
	1	28
Dimer+Tzb	1	4
	1	38

[a] Scaled time.

for the bilayer forming element. Other more complex membrane compositions may exist in the ErbB2 system but these are not clearly defined. Systems consisting of 1041 to 1052 DPPC molecules solvated in 79453 to 80642 Martini water particles were prepared to build the different lipid bilayers. A triclinic box with cartesian dimensions 17x17x35 nm was selected so that periodic image interaction were avoided. The protein was positioned so that the TM helix was inside of the lipid bilayer.

Several replicas were performed for each system. Table 6.1 collects that number and the simulation time for each one.

Simulation details

All CG molecular dynamics (MD) simulations were performed using the version 4.5.3 of the GROMACS simulation package[48]. An energy minimization using the steepest descent algorithm over 1000 steps was carried out for the initial structure. Then, a NPT equilibration with position restraints for all protein beads was run for 40 ns, and finally, a NPT equilibration without position restrains was run for 8 ns.

The Particle Mesh Ewald (PME) method[49] was used to calculate long-range electrostatic interactions, using a maximum grid spacing of 2.5 Å, fourth-order (cubic) interpolation for the fast Fourier transforms and a relative dielectric constant of 15. A dielectric constant of 15 is used for explicit screening to balance the increased hydration strength of many of the CG particle types[36]. The temperature was kept constant at 310 K by coupling the protein and the solvent independently to an external bath using the Berendsen algorithm with a coupling constant of 1 ps[50]. Isotropic scaling was used for the pressure (1 bar) with a coupling constant of 2.0 ps and a compressibility of $3.5e^{-5}$ bar^{-1} following the Berendsen algorithm[50]. The temperature range used in the original Martini force field parameterization for lipids was 270–330 K. The calculated transition temperature for DPPC is 295.5 K below the experimental value of 315K[36]. Therefore, the 310 K temperature selected for our simulations could

be considered physiologically relevant.

The dynamics were integrated using the velocity Verlet integrator, with a time step of 20 fs and bonds constrained using the LINCS algorithm[51]. Thus, production dynamics were performed at constant pressure and temperature (NPT ensemble) storing the trajectory every 0.5 ns to increase the statistical significance. Independent trajectories were generated for each studied model using different seed numbers for the initial velocity assignment.

It has been proposed that the time unit used in coarse grained simulations using the Martini force field should be multiplied by four to obtain an estimation of the corresponding atomistic time unit[52]. This conversion factor is derived from the diffusion dynamics of CG water compared to atomistic water. Therefore, the dynamic properties analyzed through the results section are reported in the scaled time units.

Distance maps between the backbone particles were calculated in order to analyze the interdomain interactions. Normalized maps were constructed by counting the number of frames where the inter residue backbone particle distance is below a threshold of 0.7 nm and dividing by the total number of frames. Thus, a value of one in the distance map would corresponds to a persistent distance below 0.7 nm between the corresponding residue backbone particles along the whole trajectory.

Principal component analysis (PCA)[53, 54] is a method that takes the trajectory of long MD simulations and calculates the dominant modes in the molecular motion. Thus, the configurational space is reduced, containing few relevant collective degrees of freedom in which long range fluctuation can be studied. A PCA diagonalizes the covariance matrix of the atom fluctuations from their average trajectory. In this framework, the larger eigenvalues correspond to eigenvectors which explain most of the variance associated to particle fluctuations. The eigenvalue ordering allows to identify a small set of modes that capture most of the fluctuations. We have performed a PCA analysis in order to identify the lowest frequency motions occurring in the tyrosin-kinase dimer. The g_covar and g_anaeig tools in the GROMACS package were used to perform the PCA analysis.

6.3 Findings for the Monomer and Homodimer Systems

CG simulations were performed on both the monomer and dimer forms of the ErbB2 receptor in a bilayer model. The complexes with the Tzb-Fab bound to its corresponding epitope at the ECD are modeled in order to study the antibody effect on the structure and dynamics of the receptor. Figure 6.1 depicts both the Trastuzumab Fab antibody bound to the ECD of the monomer and dimer forms of the full ErbB2 receptor embedded in a lipidic bilayer membrane model.

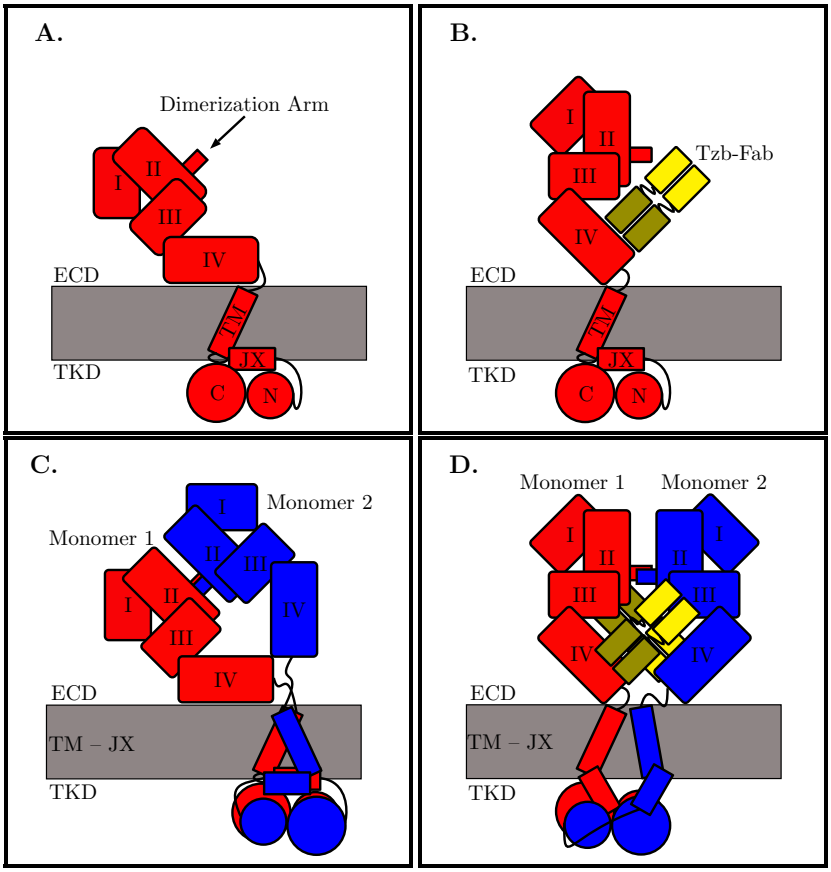


Figure 6.1: Schematic Representations. A Antibody-Free ErbB2 monomer receptor. The subdomains I, II, III and IV of the ECD domain, TM domain, JX domain and the TKD domain are shown in red. The C and N label stands for the C-lobe and N-lobe domains of the TKD. The shadow area represents the lipid bilayer. **B Antibody-Bound ErbB2 Monomer** receptor. The Tzb-fab is bound to the subdomain IV of the ErbB2 receptor. The variable (VD) and constant (CD) domains are shown in green and yellow colors, respectively. **C Antibody-Free ErbB2 Homodimer**. **D Antibody-Bound ErbB2 Homodimer**, the two antibodies bound to their respective subdomains IV are shown with the same color code.

In what follows, the structure and dynamics of the ECD, TM and TKD domains will be separately discussed for both the monomeric and dimeric forms without and with Trastuzumab bound to the ECD, respectively. A comparison with other available atomistic simulation models as well as with experimental data will be done, whichever is possible, in order to evaluate the performance of the CG models. Finally, we make a discussion of all results.

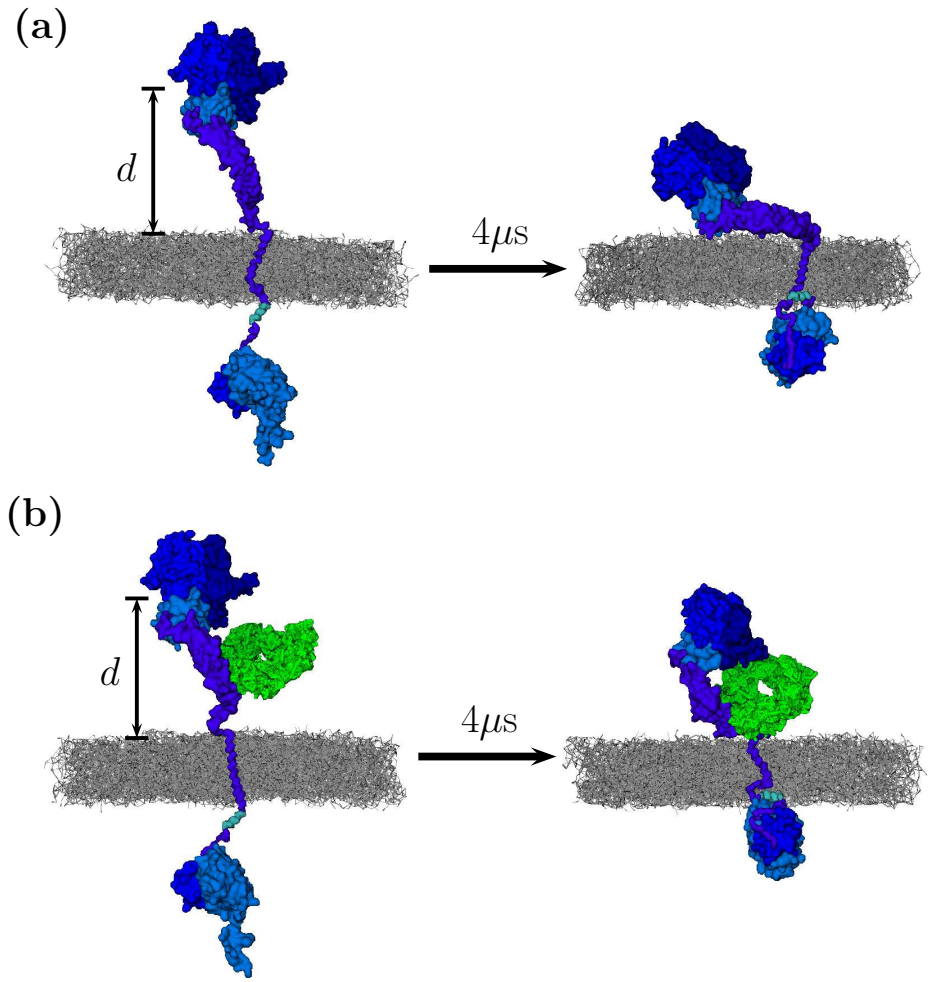


Figure 6.2: Initial and final conformations of a selected replica of **(a)** the antibody free ErbB2 monomer and **(b)** the antibody-bound ErbB2 monomer. ErbB2 receptor and antibody are shown in blue and green, respectively. Space fill representation for the CG particles. Lipid bilayer molecules are depicted as gray lines and water particles are omitted. The distance d is defined to measure the approach of the receptor to the lipid bilayer (see text for details). The initial conformations of the ErbB2 monomer were prepared using the Bagossi's structure

6.4 ECD domain

6.4.1 Monomeric form without and with Trastuzumab

Selected initial and final snapshots of the monomer simulations without (antibody-free) and with (antibody-bound) Tzb are depicted in Figure 6.2. Firstly, in the antibody-free monomer system the ECD presents a hinge movement at the inter-

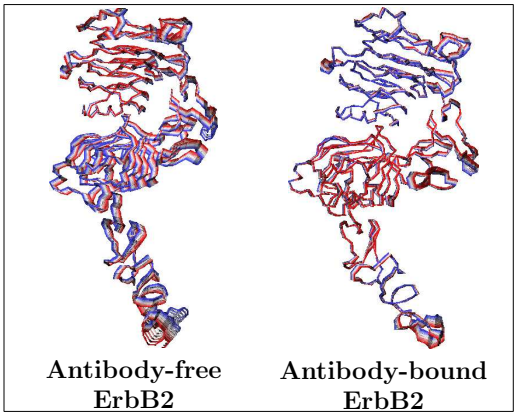


Figure 6.3: First PCA eigenvector for the ECD-ErbB2 monomer corresponding to the antibody-free (left) and antibody-bound (right) systems. Only the backbone is represented (as connecting lines). Several projections of the backbone are plotted between the two extremes denoted with red and blue colors.

face of subdomains III and IV (residues 460–480) that approaches subdomain IV to the dimerization arm (subdomain II). Figure 6.3 shows the first eigenvector of the PCA analysis performed on simulations of both monomeric systems. A remarkable similarity can be observed to the result obtained with the PCA analysis performed on atomistic simulations of ECD-ErbB2 (see, for example, Figure 4.5). This movement in the CG simulations agrees also with previous atomistic simulations performed on the monomeric ErbB2 extracellular domain[41, 47].

Remarkably, a similar motion has been recently described by MD simulations of the full EGFR receptors with and without ligand[34]. This structure has not been yet reported by crystallographic techniques likely due to the rigid packing of the produced crystals[41]. Thus, the CG model reproduces the intrinsic flexibility of the ECD domain observed in atomistic simulations. Furthermore, the presence of the membrane does not seem to hinder this intrinsic flexibility of the ECD.

A closer comparison with the atomistic simulations can be seen in Figure 6.4. This shows the RMSF analysis of the residues in the ECD domain for both the CG and atomistic simulations of the monomer taking as a reference the Bagossi's structure [33]. The plot compares the fluctuation values for the C_α in the atomistic models against the backbone particles of the CG particles (mapped in the C_α atoms in Martini). In spite of the elastic networks treatment of the CG model (see details in Computational Methods section), which is expected to reduce fluctuation values, the profile is rather similar in both cases, showing higher values in the loops and in the subdomains involved in the hinge motion above mentioned.

The binding region between the ECD-ErbB2 epitope (subdomain IV) and the Variable Domains of both the Heavy (V_H) and the Light Chain (V_L) of the monoclonal antibody has been experimentally characterized by three ErbB2 loops. Loops formed by residues 557–561 and 593–603 are mainly electrostatic, whereas the 570–573 loop makes principally hydrophobic contacts with the V_H of the antibody. Figure 6.5 shows the distance maps of the contact interface between

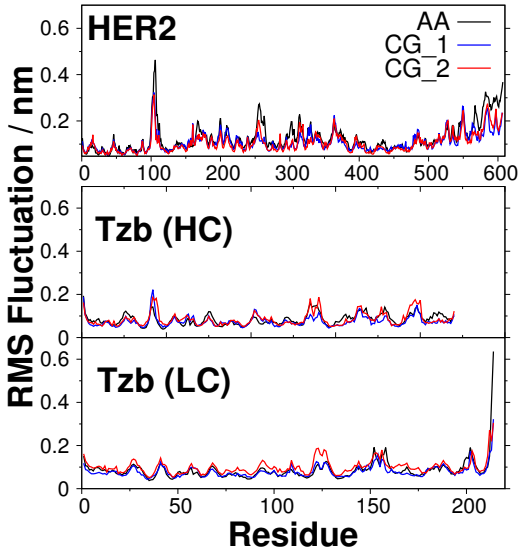


Figure 6.4: RMSF plots for the ECD-ErbB2 (top), Tzb-Fab heavy chain (HC, middle) and light chain (LC, bottom). The atomistic simulations were performed on the antibody-bound ErbB2 ECD in water. The CG_1 and CG_2 labels correspond to two independent CG MD simulations. The reference structure in all cases was taken from the Bagossi's structure

the subdomain IV and the antibody variable domain for the crystal, atomistic and the CG structures, respectively. As it can be seen, the variable domain of the monoclonal antibody structure maintain its interactions with the ECD-ErbB2 epitope (subdomain IV) in all CG replicas as compared to the experimental structure (PDB code: 1N8Z)[1]. The same observations have been obtained in the second replica as can be seen in distance matrix showed in the Figure 6.17. Furthermore, a hinge movement is also observed in this case, which favors the interaction of the antibody constant domain with the dimerization arm in subdomain II. Thus, residues Thr256–Pro263 at the dimerization arm are found at distances below 0.7 nm from Fab residues Ser120–Gly125 and Ser210–Asn211 in the Heavy Chain (HC), in line with the interactions reported for the atomistic simulations[41].

Along the manuscript, the comparison with the atomistic model was described using the ECD from the full-receptor structure (Figure 6.4 and 6.5). A most appropriate comparison should be made with the CG model of the ECD-ErbB2/Tzb-fab complex alone. The comparison of RMSF plots that was discussed in answering the EN question is a source of validation of the coarse-grained model. Another source of validation regarding the protein–protein interactions are the distance matrices evaluated between the ECD epitope and its interaction with the Tzb-Fab. In Figure 6.18, the distance matrices calculated on the atomistic and coarse-gained models for the ECD and Trastuzumab complex are shown. It can be observed that most of the crystallographic contacts are well preserved in the atomistic as well as in the Martini system.

A third validation test contemplates the dynamic properties of the system. A comparative PCA analysis of the atomistic and CG simulations following the

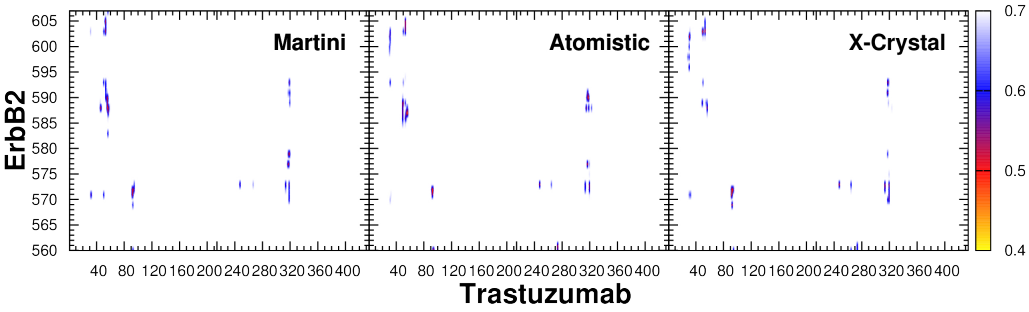


Figure 6.5: Distance matrix between residue pairs of the ErbB2 subdomain IV and the variable domain (VD) of the antibody monomer. The distance corresponds to the smallest distance between the CG-beads of two residues in the receptor and in the antibody, respectively. A cut-off distance of 0.7 nm was used. The experimental structure (PDB code: 1N8Z) was mapped on the MARTINI CG model

method proposed by Siuda and Thogersen[55] was carried out. The eigenvectors found when diagonalizing the covariance matrix were then used in the root mean-square inner-product (RMSIP) analysis (see equation 1 of [55]), that quantifies the overlap between the essential subspaces (described by the 10 first eigenvectors) obtained from the AA and CG simulations. The RMSIP obtained was 0.51, which may be considered satisfactory taking into account that a value of 1 means a perfect overlap and 0 refers to an orthogonal space.

In Figure 6.6, it is shown two snapshots corresponding to a fitting between the backbone CG particles and the Cα atoms of the atomistic model for the ECD-ErbB2/Tzb-Fab complex. A RMSD value of 0.8 nm was obtained for the full complex, whereas the RMSD values calculated separately for each constituent were 0.4 nm for the ECD chain and 0.3 for the Tzb-Fab. In this figure, it can be also observed the additional interaction between the antibody constant domain and the ErbB2 dimerization arm that has been captured similarly by both atomistic and CG simulations.

As it can be seen in Figure 6.2, the final configurations of the protein with respect to the lipid bilayer are rather different in the free and bound-antibody monomer forms. In all replicas of the free-antibody monomer, an approach of the protein to the lipid bilayer surface is observed. On the contrary, this approach has not been detected in the case of bound-antibody complexes. In our simulations, the vertical distance between the lipid polar head particles and the backbone center of mass (COM) corresponding to subdomains I, II and III was evaluated to quantify the approach of the ectodomain to the bilayer surface (**d** in Figure 6.2). Indeed, in the monomer systems, the calculated distance **d** drops from the initial 10.5 nm to final values around 4.0 $\pm$ 0.3 nm and 7.9 $\pm$ 0.3 nm for the free and the bound-antibody monomer forms, respectively.

The approach of the free-bound antibody monomer is in agreement with some

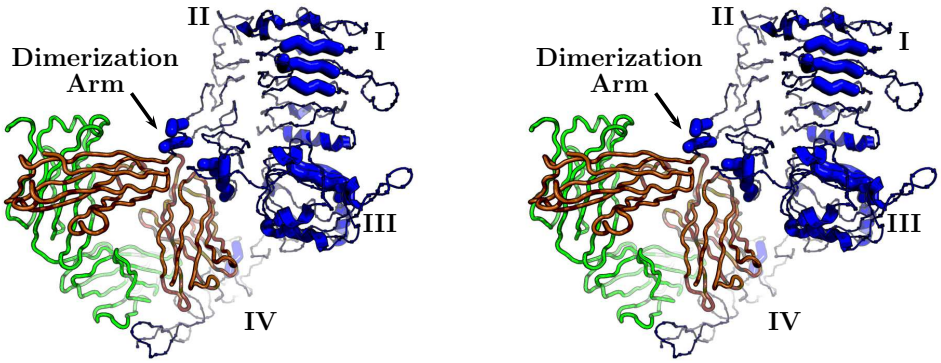


Figure 6.6: ECD-ErbB2/Tzb-Fab complex showing the interaction between the dimerization arm and the antibody constant domain. Left: atomistic model. Right: CG model. The two Fab chains are depicted in green and brown colors. Roman numbers indicate the location of the subdomains in the ErbB2 receptor. The ECD-ErbB2/Tzb-Fab (CG) was extracted from the full receptor simulation

FRET experiments for the homolog EGF-ErbB1 homodimer where no antibody is present[31, 32, 56]. In those experiments, it has been shown that the ErbB1 ectodomain lies parallel to the cell membrane at distances around 4.0 nm. The geometric parameter used in the FRET experiments for the homolog ErbB1 (EGFR) receptor is the distance between a labeled Epidermal Growth Factor (EGF) ligand bound to the ErbB1 receptor and the bilayer surface, which is comparable to the **d** distance for our system defined above. These authors have proposed that this approach of the receptor to the membrane is caused by the high flexibility of the subdomain IV in the ECD. Anyhow, the antibody bound to the subdomain IV is also anchored to the bilayer membranes (see final state in Figure 6.2b), which in our models does not allow the approach of the receptor to the membrane in the case of the bound-antibody monomer.

The final structures corresponding to the antibody-free monomer present persistent interactions between residues 359–364 and 393–397 of the subdomain III, residues 481–484 and 611–620 of subdomain IV and the bilayer surface. These fragments contain several negatively charged ASP and GLU amino acids able to interact with the positively charged lipid choline heads of the bilayer. In the antibody-bound monomer case, the Tzb is able to develop permanent interactions with the bilayer surface, as above mentioned, hindering the approach process of the receptor to the lipid bilayer. The closest residues of the antibody to the bi-

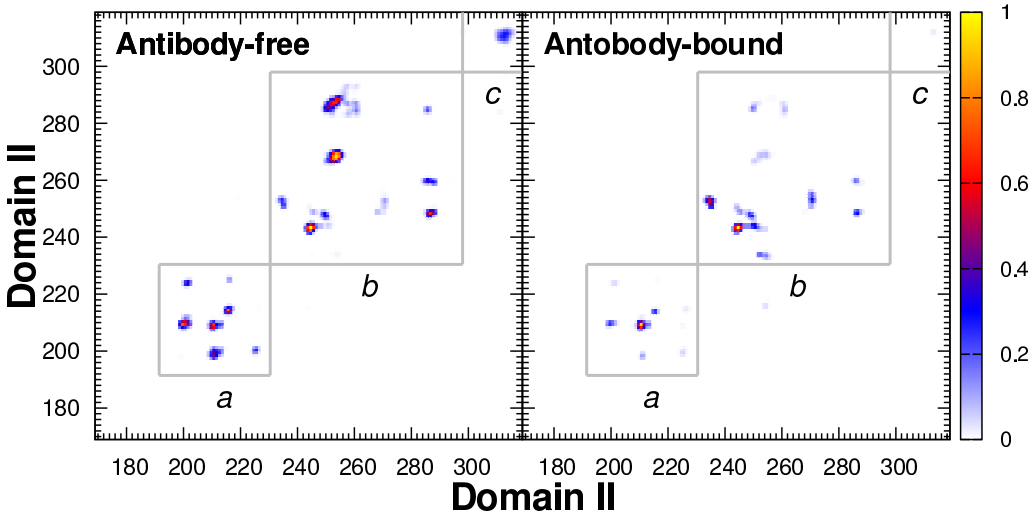


Figure 6.7: Normalized contact maps for subdomain II – subdomain II interaction between both monomers in the antibody-free. **(A)** and the antibody-bound **(B)** ErbB2 dimers. This map is constructed by counting the number of frames where the inter residue backbone particles distance is below a threshold of 0.9 nm and dividing by the total number of frames. Lowercase letters stand for the different regions of contact between dimerization arms. The color scale is in the range of 0 (no interaction along the trajectory) to 1 (full interaction along the trajectory)

layer surface are Gly16, Asp17, Arg18 and Ser76 located at the heavy chain of the Fab.

Several replicas of the simulated systems were performed as stated in Table 6.1. As it might be expected all replicas do not behave identically. A brief discussion about variance across replicas has been added as MD monomer replicas discussion in section 6.9 (*Monomer MD Simulations: Replicas*).

6.4.2 Dimeric form without and with Trastuzumab

The antibody-free ECD dimer structure (Figure 6.1C) is mainly driven by the interaction between the dimerization arms at the subdomain II of each monomer. The overall interaction through the dimerization arms originally found in the crystallographic structures of the ECD-ErbB1[57] and subsequently used by Bagossi for his ErbB2 model[33] is preserved along the CG simulations. However, some weakening of the dimer interaction appears in the antibody-bound dimer model (Figure 6.1D). Figure 6.7 presents the normalized residue contact maps for subdomain II – subdomain II complex interface for the antibody-free and antibody-bound models, respectively. It can be observed that the residue contacts in regions A and B are weakened in the antibody-bound dimer model. Furthermore, the interaction in the subdomain II C-terminal region (residues 306–309 in

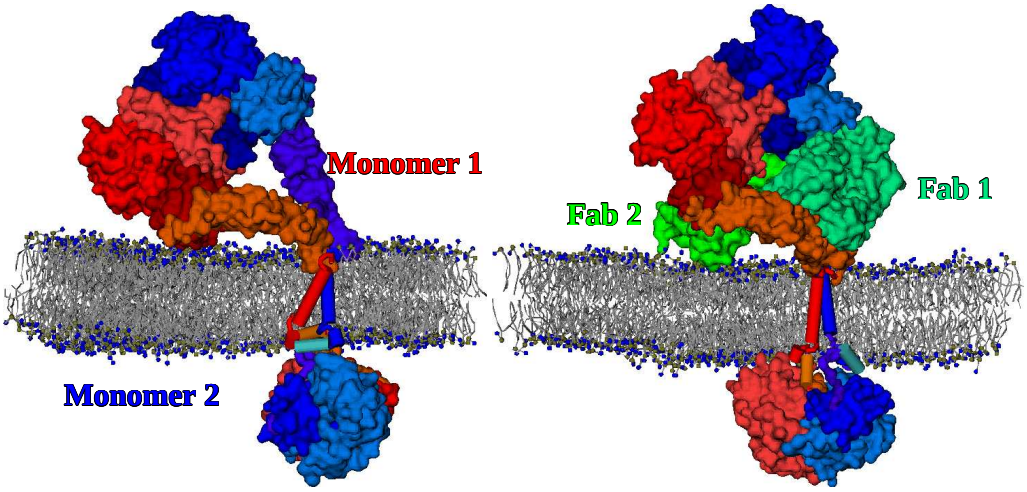


Figure 6.8: Full antibody-free (left) and antibody-bound (right) ErbB2 homodimer structures at the end of the simulation. Space fill representation of the dimeric structures of ErbB2 are shown in red (monomer 1) and blue (monomer 2) for each monomer. On the other hand, the antibody (Trastuzumab) is represented in green. Lipid bilayer molecules are depicted as gray lines. Water particles are omitted for clarity

region c) has disappeared in the antibody-bound case. Thus, the proximity of the constant domain of the antibody disturbs the loops at subdomain II connecting the dimeric interface. The variable domain of the antibody structure maintains its interaction with the ErbB2 extracellular epitope (subdomain IV). The Tzb – subdomain IV interaction observed in the monomer is rather similar to that found in chain 1 of the ErbB2 dimer(see Figure 6.8). The other ErbB2 chain presents a somewhat different interaction pattern with Trastuzumab. Figure 6.19 in section 6.9 shows the distance matrices for the two monomers corresponding to the longest replica of the Trastuzumab-bound homodimer system. The matrix labeled "Monomer 1" correspond to the monomer lying on the bilayer surface (see Figure 6.8). This matrix is essentially similar to that described for the monomer case (see Figure 6.5) and comparable to the crystallographic structure. The matrix labeled "Monomer 2" corresponds to the monomer standing upright. The residues involved in the ECD epitope – Tzb-Fab interaction are maintained. However a different distance matrix pattern is obtained concerning residues 560–565 and 586–593 of the ErbB2 chain. These interactions may correlate with the loose interface in the subdomain II dimerization region for the Trastuzumab-bound homodimer system.

The C-terminal fragment of subdomain IV is another structural characteristic of the ECD dimer structure. Atomistic scale simulations performed on ErbB2-ErbB3 ECD dimers in solution show a clear interaction between the C-terminal segments of the subdomains IV [42]. This is also the situation exhibited by some

crystallographic structures of homolog receptors, e.g., the ECD-ErbB1 dimer[57]. The CG simulations of the antibody-free dimer system show the expected subdomain IV C-terminal interaction between the two ErbB2 monomers. The contact map associated to subdomains IV for both monomers presents a remarkable interaction between the C-terminal fragments in the antibody-free system. These contacts disappear in the antibody-bound dimer complex (see Figure 6.20). Thus, the antibody bound in the vicinity of subdomain IV exerts a hindering effect on the association of these subdomains at their respective C-terminal fragments located above the membrane surface.

In Figure 6.8, the final conformation of the antibody-free (on the left) and antibody-bound (on the right) homodimer systems are shown. It can be observed that for the antibody-free homodimer, one of the ECD-ErbB2 monomer lies on the bilayer surface (monomer 1) whereas the other monomer remains in a more upright position (monomer 2). The measured distances from the ECD COM to the bilayer surface, as defined in the previous section, reach a plateau with values of 7.8 ± 0.3 and 4.0 ± 0.3 nm for monomer 2 and 1, respectively. These values agree with the FRET measurements reported on the ErbB1 receptor [31, 32, 56] and shown in the previous section. In the system containing Trastuzumab, the presence of the antibody hinders in some way the approach of any of the ECD-ErbB2 segments to the membrane (see Figure 6.8 left). In analogy to what is observed in the antibody free case, one monomer approaches the bilayer surface more than the other, keeping distances from both ECD to the lipid layer of 5.5 and 8.5 nm for the labeled monomer 1 and 2 in Figure 6.8 left, respectively. These distances are larger than the corresponding to the antibody-free system due to the interposition of the antibody between the bilayer and the ErbB2 receptors.

A comparative analysis of these results with atomistic models reported for the ErbB1 dimer can be done. The ECD-ErbB1 dimer model described in references[30, 34] for the active forms is very similar to our CG model. The interaction between the C-terminal fragments of both subdomains IV is a weakly crossing interface so that the N-terminal portions of the transmembrane domains can be close. The inactive form described in reference [34] shows both C-terminal fragments far enough to avoid the N-terminal transmembrane interaction. The Trastuzumab effect is to weaken the subdomain IV interaction, giving rise to a distortion of the transmembrane domain association. The relative orientation of the ECD domains to the bilayer surface may be driven by the electrostatic properties of the lipid polar heads. In reference [30] the lipid molecules are zwitterionic and the active ECD-ErbB1 domains lay down the lipid bilayer. The same ECD domains remain in an upright position when a mixture of zwitterionic and charged lipids is considered (reference [34]). Our CG models, which takes into account zwitterionic lipids, shows a similar behavior to the active ErbB1 system described in reference [30]. We also found one monomer being closer to the bilayer surface than the other. A more detailed discussion about the relative disposition

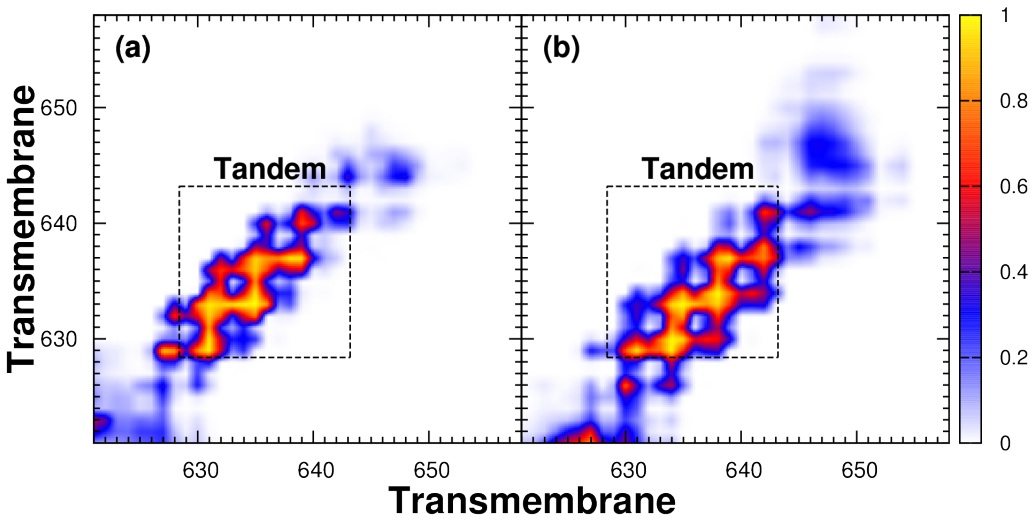


Figure 6.9: Normalized contact maps of the transmembrane (TM) interaction for the antibody-free **(a)** and antibody-bound **(b)** ErbB2 homodimer. The scale represents the occupancy of the interaction along the analyzed trajectory. See computational section for details of how the contact map is calculated. The tandem makes reference to the GxxxG-like motifs along the chain

of the ECD domains respect to the lipid bilayer has been provided above.

6.5 Transmembrane domain

The dimeric association of transmembrane domains has been related to have a key role in the activation of the entire receptor [9, 58]. In this section we describe the results obtained on the transmembrane association of the homodimer systems from CG simulations.

The starting structures for the simulations, which come from the Bagossi's model, present the transmembrane fragments interacting mainly through the N-terminal and central sections of each chain, namely, residues Arg625–Leu640. This arrangement of the TM dimer is maintained along the simulations. Figure 6.9 shows contact maps for the residues belonging to the transmembrane domains collected during the last μs of the antibody-free (Figure 6.9a) and antibody-bound (Figure 6.9b) homodimers. Firstly, the contact maps of the antibody-free homodimer are in agreement with the TM helix packing interface reported for the ErbB2 dimer by Bocharov based on NMR experiments (PBD code: 2JWA)[9]. These authors hypothesized that the arrangement experimentally elucidated would correspond to an active form of the TM dimer. The dimer is favored by the interaction between a tandem variant of GxxxG-like motifs located at the N-terminal section (see Figure 2F of reference [9]). Our CG simulations arrive to a homolog in-

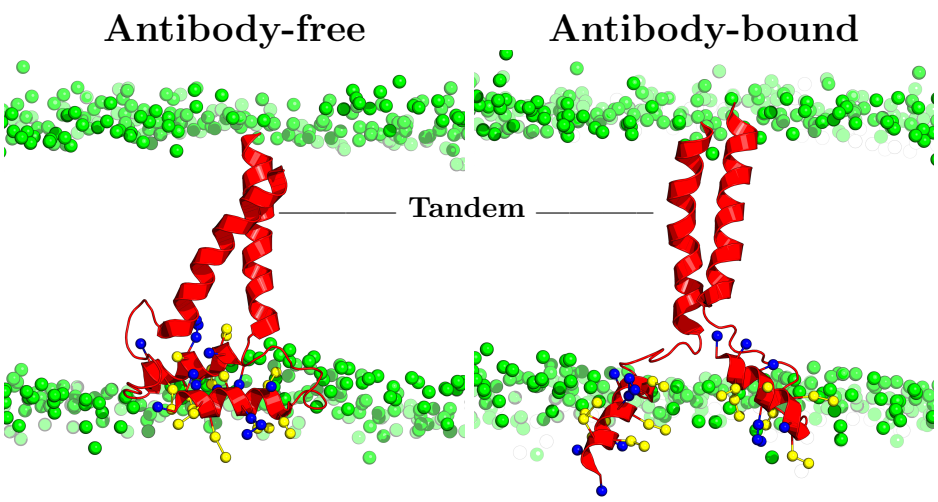


Figure 6.10: Cartoon representation of the transmembrane (long α-helix) and juxtamembrane (short α-helix) fragments extracted from the final snapshot of the antibody-free and antibody-bound ErbB2 dimer simulations. The lateral chains of the juxtamembrane domain are shown as blue and yellow CPK balls for the charged and hydrophobic residues, respectively. Lipid polar head particles are represented as green CPK balls. Water and the most lipid atoms are omitted to gain clarity in the figure.

teraction tandem corresponding to residues Ser627-xxx-Phe631-xxx-Thr635-xxx-Val639 (see Figure 6.9). In addition, The contact maps are also very similar to those reported by Arkhipov et al. for the homolog EGFR dimer in the disposition associated to an active form of the complex (see Figure 4 of reference [34]).

On the other hand, the system containing the antibody develops additional interactions between amino acids located close to the C-term region of the TM, which are not seen for the antibody-free case (please compare Figure 6.9a and Figure 6.9b). It can also be observed that in the antibody containing system, a transmembrane helix is displaced asymmetrically with respect to the other one (see Figure 6.10). This unsymmetrical arrangement is depicted in the contact map by the off-diagonal marks near the N-terminal section (residues Pro620 and Ala626, see Figure 6.9b). It can also be observed a four residue displacement of the interaction tandem towards the TM C-terminal part (Thr630-X3-Ala634-X3-Gly638-X3-Phe642 in Figure 6.9).

6.6 Juxtamembrane domain

The juxtamembrane domain (JX) (residues Lys659–Leu669), which is located between the transmembrane and kinase domains, is related to important regulatory functions such as regulation of the receptor trafficking and inactivation[59].

Table 6.2: Geometrical parameters of the juxtamembrane arrangement. The label n.a means not applicable

System	**Torsion** $\varphi(^o)$	**Angle** $\theta(^o)$	
		Monomer 1	*Monomer 2*
Antibody-free monomer	n.a	19 ± 2	n.a
Antibody-bound monomer	n.a	21 ± 2	n.a
Antibody-free dimer	7 ± 5	7 ± 3	3 ± 2
Antibody-bound dimer	57 ± 9	27 ± 4	60 ± 3

Thus, the arrangement of this JX domain with respect to the bilayer and to the TKD has been studied and related to regulatory functions of the EGFR family.

In this section, we define the torsion angle (φ) between the JX α-helix axes as well as the θ angles between these axes and the bilayer surface, *i.e.*, θ gives an idea of the position of the JX domain respect to the lipid bilayer whereas the φ shows the relative orientation of the JX domains in the case of the homodimer complex. These definitions are shown in Figure 6.11. They help to quantify the JX spatial arrangement.

6.6.1 Monomeric form without and with Trastuzumab

In the monomer case, it can be observed that the juxtamembrane segment adopts a parallel configuration with respect to the bilayer surface so that the polar residues are oriented to the water phase whereas their hydrophobic amino acids face the lipid interior. This observation agrees with the NMR experimental results obtained on the EGFR juxtamembrane fragment[60]. The angle (θ) formed by the juxtamembrane helix axis and the lipid bilayer surface is similar for both cases (19 ± 2 for the Antibody-Free ErbB2 and 21 ± 2 for the Antibody-bound complex).

6.6.2 Dimeric form with and without Trastuzumab

The Antibody-free homodimer reach an antiparallel arrangement of both juxtamembrane α-helices ($\varphi \sim 7^o$, see Table 6.2) as well as a parallel disposition with respect to the bilayer surface ($\theta \sim 3^o$ and 7^o, see Table 6.2). It can be noted that the antiparallel disposition has been identified to be essential for an active dimer conformation[34, 58].

On the other hand, it can be observed a severe distortion of the juxtamembrane antiparallel arrangement in the antibody-bound homodimer system ($\varphi \sim 57^o$, see Table 6.2). In the same way, the parallel alignment respect to the bilayer surface is lost ($\theta \sim 27^o$ and 60^o, see Table 6.2).

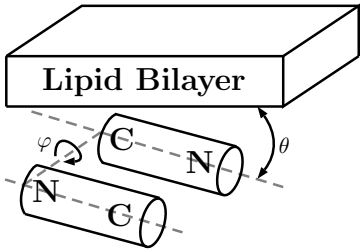


Figure 6.11: Schematic representation of the mutual disposition of the JX fragments (cylinders) and the lipid bilayer (prism).

Besides the parallelism with the bilayer exhibited by the juxtamembrane segments in the antibody-free simulations, these fragments remain anchored on the surface of the membrane, as can be observed in Figure 6.10. The hydrophobic residues Ile660, Tyr663, Met665, Leu668 and Leu669 (yellow CPK balls in Figure 6.10) face the lipid acyl tails whereas the Lys and Arg charged amino acids (blue CPK balls in Figure 6.10) interact with the lipid polar heads. This picture of the interaction between the juxtamembrane fragment and the lipid bilayer are in agreement with the NMR results [59] and with the description recently provided for the EGFR dimer[34, 58] in its active conformation. In the case of the antibody-bound homodimer almost all these interactions are weakened, losing the antiparallel arrangement of both juxtamembrane α-helices.

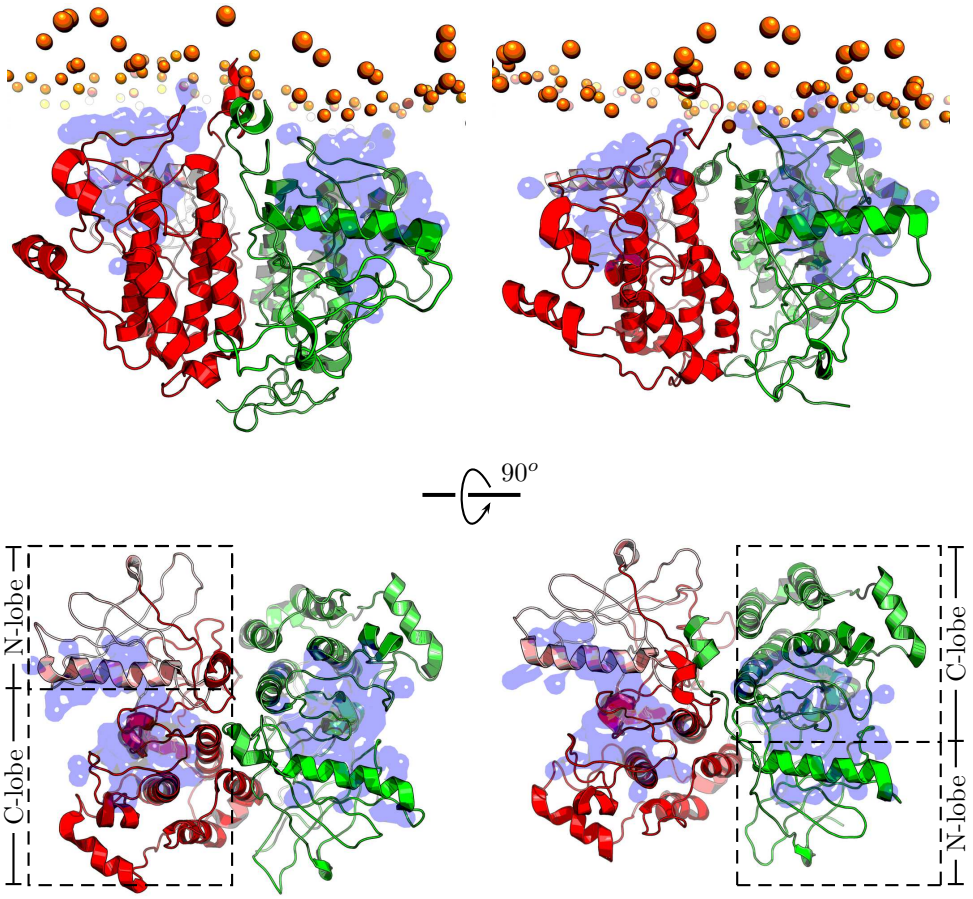
6.7 TKD intracellular domain

Regarding the TKD, the general observation in all the systems is that the TKD approaches the lipid bilayer surface on the intracellular side.

In the monomer cases, the distance from the tyrosin-kinase COM to the bilayer surface is 2.6 $\pm$ 0.2 nm for both systems. The tyrosin-kinase residues at distances to the membrane surface smaller than 0.8 nm correspond to the residues Ile850–Gly859, Leu873–Arg875 and Asp940–Cys943.

The representation for the dimer structures in Figure 6.8 apparently shows a closer approach of the TKD dimer to the bilayer surface in the case of the antibody-bound ErbB2 homodimer. Looking at the distances between the TKD-COM and the bilayer surface, averaged over the last μs of all the corresponding replicas, the cytoplasmic domains of the antibody containing system present a slightly closer approach to the membrane surface (2.8 $\pm$ 0.2nm) than the antibody-free dimer case (3.0 $\pm$ 0.2nm). In Figure 6.21 the distance between the TKD-COM and the bilayer surface along the simulation time for several replicas of the Trastuzumab-free homodimer system is shown. The approach to the bilayer surface occurs in the first microsecond reaching a distance between both TKDs COM to the bilayer surface that fluctuates between 2.5 and 3 nm, as illustrated in that Figure.

Figure 6.12 depicts the disposition of the TKD dimer relative to the bilayer



Antibody-free ErbB2 dimer **Antibody-bound ErbB2 dimer**

Figure 6.12: Cartoon representation of the tyrosin-kinase domains extracted from the final snapshot of the antibody-free and antibody-bound ErbB2 dimer simulations. Upper: Relative position of the TKD dimer respect to the bilayer. Lipid polar head particles are represented as green balls. Lower: TKD dimer image taken from the normal to the bilayer viewpoint. Lipid particles are omitted. The N and C lobes are shown for one of the chains to illustrate the back-to-back arrangement. Magenta shadow surfaces correspond to residues involved in the active site

surface. There are three groups of residues that permanently approach the bilayer surface at distances below 0.8 nm, namely, Gly850–Lys859, Leu873–Arg875 and Trp937–Cys943 located in the C-lobe of the kinase domain. The first and third fragments contain charged residues able to interact with the lipid polar head particles. This result is in line with the X-ray elucidated structure of the TKD ErbB1 dimer complexed with a lipid-like molecule[59]. It can also be observed in

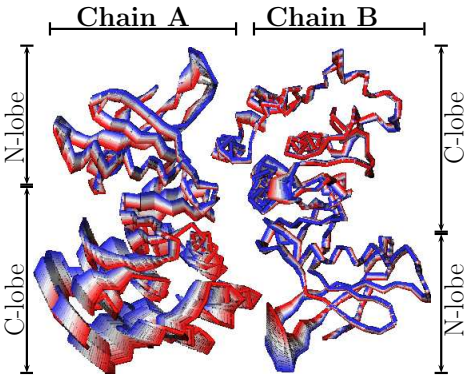


Figure 6.13: First PCA eigenvector for the tyrosin-kinase dimer. Only the backbone is represented (as connecting lines). Several projections of the backbone are plotted between the two extremes denoted with red and blue colors

each case that the TKD dimer approaches the bilayer asymmetrically, being one monomer closer to the bilayer surface. This result is in close agreement with the full atomistic models of the EGFR dimer in a lipid bilayer[34].

The interaction between both tyrosin-kinase domains corresponds to the back-to-back configuration as was proposed in the original Bagossi representation. The N-lobe of one chain interacts with the C-lobe of the other, leaving the active site in the opposite sides of the contact surface formed by the $N \rightarrow C \rightarrow C \rightarrow N$ interaction of both TKDs (see Figure 6.12). As the dynamics evolves, the back to back interaction is continuously lost due to a progressive disruption of one of the two N-lobe C-lobe interactions. In order to examine the opening movement of the tyrosin-kinase dimer, we performed a Principal Components Analysis (PCA) on the last μs of the trajectory focusing on the backbone particles of the tyrosin-kinase domain. The first eigenvector extracted from the fluctuation analysis shows the increased displacement of one tyrosin-kinase C-lobe with respect to the N-lobe belonging to the other ErbB2 tyrosin-kinase chain. Figure 6.13 depicts a schematic representation of the main PCA component corresponding to the Trastuzumab-free system. Several projections of the backbone are shown in different colors between the two extreme points (blue and red colors) to illustrate the displacement of one C-lobe out of its N-lobe partner.

The influence of the Trastuzumab Fab in the TKD dimer interface seems to be rather subtle as can be deduced from the simulations reported in this work. Figure 6.14 presents the time evolution of the COM distance between the N and C lobes of the two monomers (part A) and from these to the bilayer surface (part B). On one hand, one of the two interacting N-lobe C-lobe pairs in the TKD dimer is weakened along the dynamics. As can be observed in Figure 6.14, part A, the inter-COM distance between the monomer 1 C-lobe and monomer 2 N-lobe is elongated to values above 3.5 nm whereas the equivalent distance between monomer 1 N-lobe and monomer 2 C-lobe remains around 3 nm. The evolution is similar in the Trastuzumab-free and bound cases as well. This is in agreement with the PCA analysis discussed above where an analogous opening was reported.

On the other hand, the relative position of the TKD dimer respect to the bilayer surface shows some features already discussed in the monomer case. Each TKD chain shows one lobe closer to the bilayer than the other. Either the N or C lobe can alternate the approaching to the bilayer surface (see for example the red and green lines in part B graphs of the Figure 6.14). The inter-chain N lobe C lobe interaction is revealed by the superposition of red (N lobe monomer 1) and cyan (C lobe monomer 2) curves on one side and blue(N lobe monomer 2) and green (C lobe monomer 1) lines on the other. The interaction pair formed by the N lobe in monomer 2 and the C lobe in monomer 1 seems to be systematically farther from the membrane surface than the other N-lobe – C-lobe interaction. This observation correlates with the distance elongation shown by monomer 1 C-lobe and monomer 2 N-lobe interaction. Finally, a subtle observation regarding the effect of the antibody can be the slightly lower overall distance between the TKD dimer to the bilayer surface as can be deduced from the plots in part B (points in the antibody-bound case plot reach distance values lower than 2 nm more frequently than in the antibody-free case)

The TKD is marginally modified respect to the initial model proposed by Bagossi. Due to the Elastic Networks treatment, no conformational changes in the active site are contemplated and a comparison with simulations of the TKD cannot be performed at the moment. The TKD dimerization is stable in its initial conformation although some trend to an asymmetric disposition of the C–N lobes can be depicted from the PCA analysis performed on the particle dynamics fluctuations. An extrapolation of this movement will presumably drive to an asymmetric configuration of the tyrosin-kinase dimer with only an N-lobe/C-lobe interaction between the two monomers.

Once the CG model is validated against available experimental data and atomistic simulations, we discuss the effect of the monoclonal antibody Trastuzumab on the receptor dimerization at a molecular level. The simulation of the full receptor allows us to suggest an action mechanism for the Trastuzumab Fab while interacting with the ErbB2 dimer as follows. The binding of the Trastuzumab Fab to its native epitope at subdomain IV hinders the interaction of both ECDs through the subdomain IV C-terminal fragment. This defective interaction causes in turn a small disruption of the transmembrane dimerization which is slightly displaced towards the C-terminal fragment. The juxtamembrane segments in the intracellular domain should be affected by this displacement so that an antiparallel arrangement cannot be observed in our simulations. This lack of an antiparallel disposition might be associated to an inactive conformation of the TKD domain. The proposed mechanism is in line with the recent biological observations regarding the effective antibody deactivation on the proliferation signaling cascade initiated by the ErbB2 dimer[45].

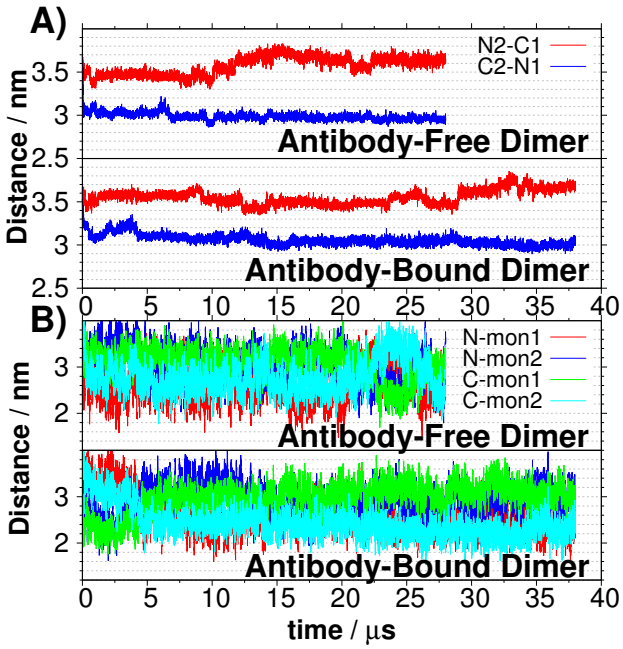


Figure 6.14: **A)** Distances between the COM of N and C lobes in the TKD dimer interface. **B)** Distances from the TKD dimer N and C lobes to the bilayer surface.

6.8 Conclusions From This Work

We report, for the first time, molecular dynamics simulations of full models of ErbB2 monomer and dimer systems embedded in a lipid bilayer and complexed with the Trastuzumab-Fab monoclonal antibody. The system has been simulated within the Coarse Grained approach using the Martini force field. This force field is a well established model for molecular dynamics simulations of systems containing proteins and lipids. The results presented in this paper support this assertion in view of the satisfactory comparison with experimental data and atomistic simulations.

Starting from the models proposed by Bagossi, we have carried out several μs coarse grained molecular dynamics simulations on systems containing either the full ErbB2 monomer or dimer with or without the monoclonal antibody Fab and embedded in a lipid bilayer. The model proposed in this work seems to fulfill a number of experimental and computational observations about the full ErbB2 structure and its interaction with the lipid bilayer. Some details observed in atomistic simulations of the different ErbB domains are reproduced in the coarse grained models. The most relevant observations are summarized below. In the monomer case an interaction between the antibody constant domain and the dimerization arm can be observed in addition to the native interaction with

the epitope in subdomain IV. This was also observed in atomistic simulations of the ECD alone. The ECD-ErbB2 domain tends to lay down on the lipid bilayer surface in agreement with recent FRET analysis and detailed atomistic simulations of homolog systems. The intracellular domain also presents a noticeable interaction with the bilayer surface. In particular, the juxtamembrane fragment has been found to partially penetrate the membrane bilayer, in agreement with NMR measurements made with this fragment in a micelle environment.

The inclusion of the Trastuzumab Fab monoclonal antibody has a pronounced effect on the structural disposition of the ectodomain, due to the hindrance exerted on those amino acids mainly responsible for the interaction at subdomain IV C-terminal. The antibody containing system shows a displacement towards the C-terminal segment of the transmembrane domain contacts in the dimer and a complete lost of the juxtamembrane antiparallel arrangement characteristic of the active state that is also observed in the pure homodimer case. This action mechanism may explain the lost of signaling ability impaired by Trastuzumab to ErbB2 homodimers experimentally found. Despite the disarrangement of the juxtamembrane segments observed in the antibody case, the TKD dimer architecture is similar in both systems probably due to the tighter interaction in the initial Bagossi's model. The PCA analysis of the different trajectories reveals a rupture of the symmetric structure driving to a more asymmetric ensemble compatible with an active version of the TKD dimer.

6.9 *Additional Information*

Protein Sequences

Table 6.3: ErbB2 Amino Acid Sequence

	Subdomain I
1-100	TQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP
	Subdomain I | Subdomain II
101-200	LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDT | NRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAG
	Subdomain II
201-300	GCARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEV
	Subdomain II | Subdomain III
301-400	TAEDGTQRCEKCSKP | CARVCYGLGMEHLREVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL
	Subdomain III | Subdomain IV
401-500	SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTA | NRPEDECVGEGLACHQLCARGHCWG
	Subdomain IV
501-600	PGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGA
	Subdomain IV | Transmembrane | Juxtamembrane | Tyrosine-Kinase
601-700	CQPCPINCTHSCVDLDDKG | CPAEQRASPVTFIIATVVGVLLFLILVVV | VGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQ | MRILKETELRK
	Tyrosine-Kinase
701-800	VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDL
	Tyrosine-Kinase
801-900	LNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKP
	Tyrosine-Kinase | Loop
901-1000	YDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVV | IQNEDLGPASPLDSTFYRSLLEDDDMGDLVDAEE
	Loop
1001-1004	YLVP

Table 6.4: Tzb-Fab Heavy Chain Sequence

	Variable Domain
1-100	EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWG
	Variable Domain | Constant Domain
101-200	GDGFYAMDYWGQGTLVTVSS | ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQT
	Constant Domain
201-220	YICNVNHKPSNTKVDKKVEP

Table 6.5: Tzb-Fab Light Chain Sequence

	Variable Domain
1-100	DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQ
	Variable Domain | Constant Domain
101-200	GTKVEIKRTV | AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHK
	Constant Domain
201-220	VYACEVTHQGLSSPVTKSFNRGEC

Monomer MD Simulations: Replicas

Regarding the monomer system, the three replicas corresponding to the trastuzumab-free case arrive to a collapsed structure of the receptor on the lipid bilayer. However, the ECD approach to the lipid surface takes different times on each replica as can be observed in Figure 6.15. In the same Figure, it can be observed that the trastuzumab-bound replicas converge to a value around 6 nm well above the average 4 nm distance in the trastuzumab-free systems. Regarding the TKD the approach is indistinguishable, reaching a plateau value of around 2.5 nm in the first 400 ns, as can be observed in the same Figure. The RMSD values corresponding to the backbone particles for all replicas were also calculated for the first 4 μs. All the replicas show that the studied systems are well stabilized arriving to similar RMSD values after the first μs. However, the antibody-bound dimer gives different RMSD values for the two replicas, both around 3 nm, lower than the other systems.

On the other hand, we include the Figure 6.17 containing the equivalent information to that provided in Figure 6.5 but for the second replica. The two Martini graphs corresponding to the different replicas show essentially the same features. The trastuzumab residues involved in the interaction are the same while the ECD interface shows minor discrepancies in residues 575–580 and 583–584.

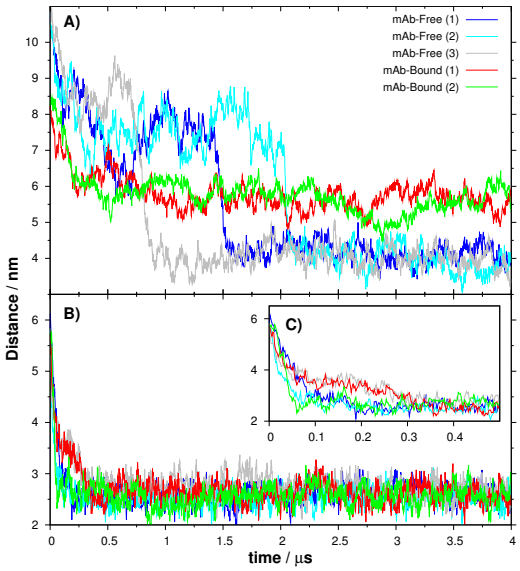


Figure 6.15: Distance from ECD-ErbB2 COM domains to the lipid bilayer surface for the different replicas described in Table 6.1 of the Chapter 6. **A)** Extracellular domain. **B)** Tyrosin kinase domain. **C)** Inset of the TKD.

R_c and K_{spring} Comparison

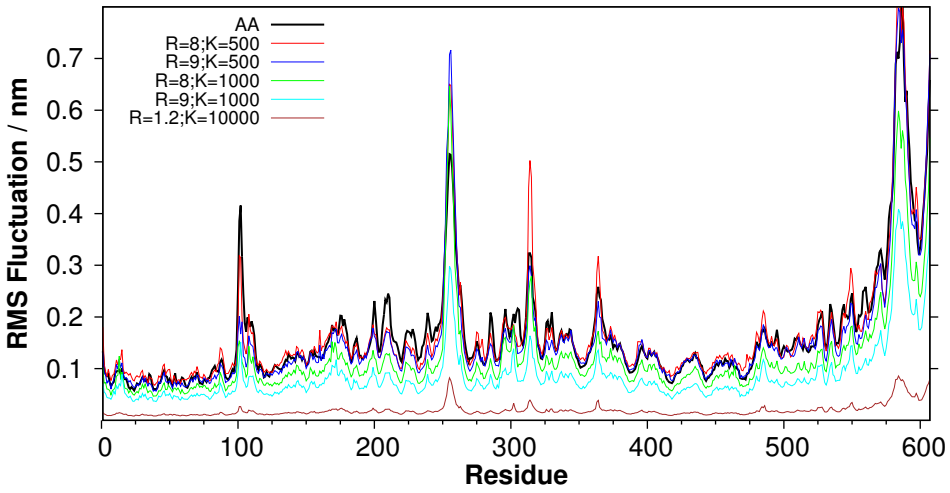


Figure 6.16: RMSF at different values of R_c and K_{spring} on ErbB2 receptor.

Distance Matrices

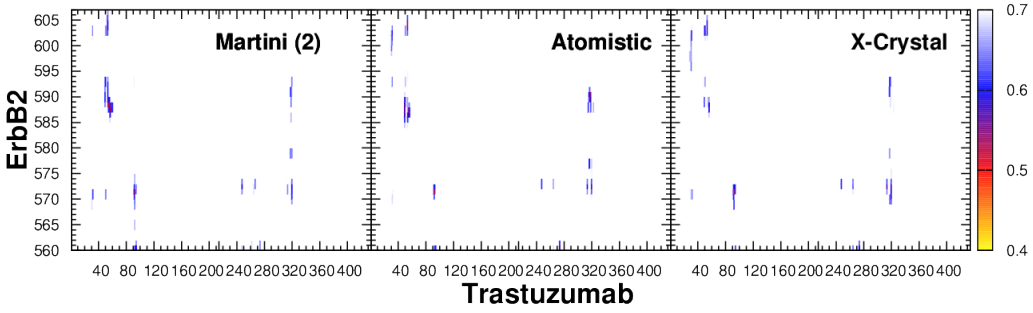


Figure 6.17: Distance matrices for the second MD replica of the trastuzumab-bound ErbB2 monomer compared with atomistic and X-ray data. The interaction between subdomain IV and the antibody Fab is plotted

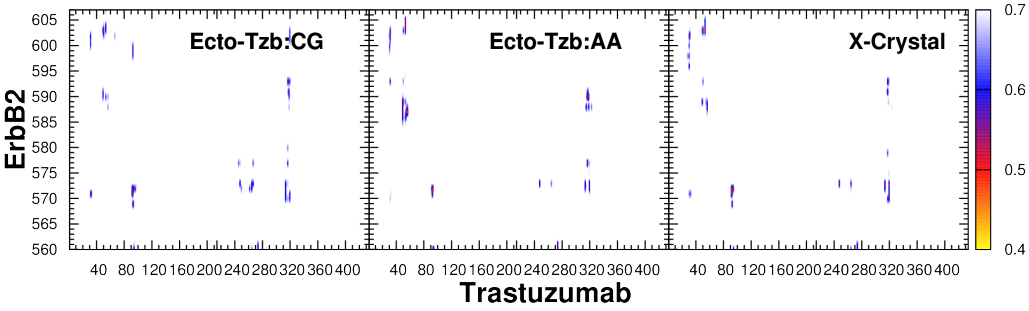


Figure 6.18: Distance matrices calculated on the atomistic and coarse-gained models for the ECD-ErbB2 and Tzb-Fab

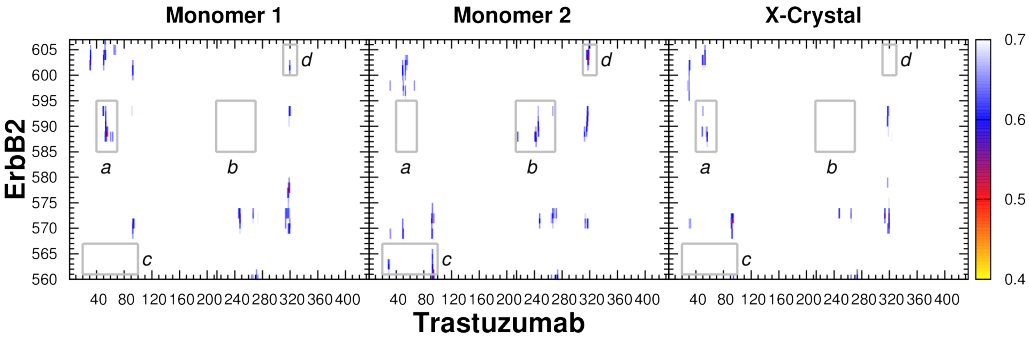


Figure 6.19: Distance matrices for the subdomain IV – Tzb-Fab interaction in the antibody-bound dimer system

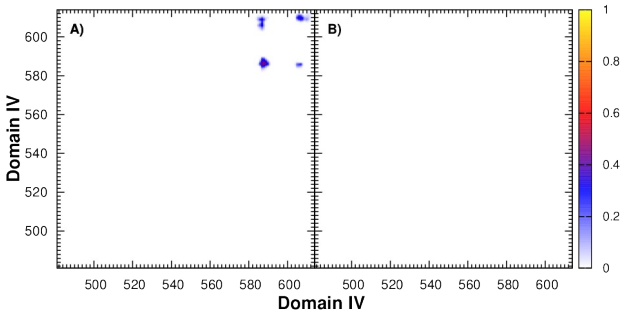


Figure 6.20: Normalized contact maps for the inter subdomain IV interaction in both dimer systems, namely, with and without Trastuzumab

TKD_{COM} – Bilayer Surface Distances

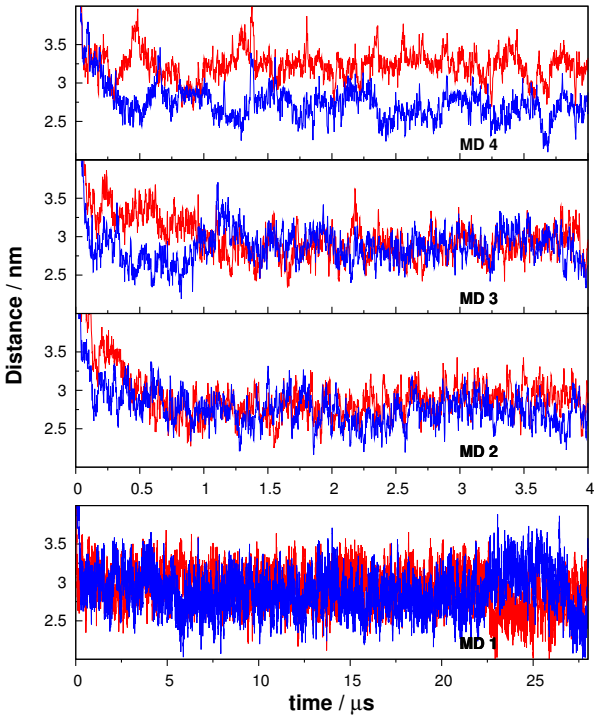


Figure 6.21: Distance from the TKD to the bilayer surface for the four replicas of the trastuzumab-free homodimer system

References

[1] Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, and Leahy DJ. *Nature*, 421(6924):756–760, 2003.

[2] Cho H-S and Leahy DJ. *Science*, 297(5585):1330–1333, 2002.

[3] Du Y, Yang H, Xu Y, Cang X, Luo C, Mao Y, Wang Y, Qin G, Luo X, and Jiang H. *J. Am. Chem. Soc.*, 134(15):6720–6731, 2012.

[4] Leahy DJ. *Cell Surface Receptors*, 68:1, 2004.

[5] Liu P, Bouyain S, Eigenbrot C, and Leahy DJ. *Protein Science*, 21(1):152, 2012.

[6] Liu P, Cleveland TE, Bouyain S, Byrne PO, Longo PA, and Leahy DJ. *Proc. Natl. Acad. Sci. USA*, 109(27):10861, 2012.

[7] Beevers AJ and Kukol A. *Journal of Molecular Biology*, 361(5):945, 2006.

[8] Bocharov EV, Mineev KS, Goncharuk MV, and Arseniev AS. *Biochimica Et Biophysica Acta-Biomembranes*, 1818(9):2158, 2012.

[9] Bocharov EV, Mineev KS, Volynsky PE, Ermolyuk YS, Tkach EN, Sobol AG, Chupin VV, Kirpichnikov MP, Efremov RG, and Arseniev AS. *J. Biol. Chem.*, 283:6950–6956, 2008.

[10] Garnier N, Genest D, and Genest M. *Biophys. Chem.*, 58:225–237, 1996.

[11] Garnier N, Genest D, Duneau JP, and Genest M. *Biopolymers*, 42(2):157, 1997.

[12] Garnier N, Crouzy S, and Genest M. *J. Biomol. Struc. Dyn.*, 21:179–199, 2003.

[13] Mineev KS, Bocharov EV, Pustovalova YE, Bocharova OV, Chupin VV, and Arseniev AS. *Journal of Molecular Biology*, 400(2):231, 2010.

[14] Mineev KS, Bocharov EV, Volynsky PE, Goncharuk MV, Tkach EN, Ermolyuk YS, Schulga AA, Chupin VV, Maslennikov IV, Efremov RG, and Arseniev AS. *Acta naturae*, 3(2):90, 2011.

[15] Prakash A, Janosi L, and Doxastakis M. *Biophysical Journal*, 99(11):3657, 2010.

[16] Soumana OS, Garnier N, and Genest M. *European Biophysics Journal with Biophysics Letters*, 36(8):1071, 2007.

[17] Samma Soumana O, Garnier N, and Genest M. *Eur. Biophys. J.*, 37:851–864, 2008.

[18] Lu P-C, Zhou C-F, Chen J, Liu P-G, Wang K-R, Mao W-J, Li H-Q, Yang Y, Xiong J, and Zhu H-L. *Bioorg. Med. Chem.*, 18(1):314, 2010.

[19] Mustafa M, Mirza A, and Kannan N. *Proteins-Structure Function and Bioinformatics*, 79(1):99, 2011.

[20] Papakyriakou A, Vourloumis D, Tzortzatou-Stathopoulou F, and Karpusas M. *Proteins-Structure Function and Bioinformatics*, 76(2):375, 2009.

[21] Qiu C, Tarrant MK, Choi SH, Sathyamurthy A, Bose R, Banjade S, Pal A, Bornmann WG, Lemmon MA, Cole PA, and Leahy DJ. *Structure*, 16(3):460, 2008.

[22] Qiu K-M, Wang H-H, Wang L-M, Luo Y, Yang X-H, Wang X-M, and Zhu H-L. *Bioorg. Med. Chem.*, 20(6):2010, 2012.

[23] Shi F, Telesco SE, Liu Y, Radhakrishnan R, and Lemmon MA. *Proc. Natl. Acad. Sci. USA*, 107(17):7692, 2010.

[24] Shih AJ, Telesco SE, Cho SH, Lemmon MA, and Radhakrishnan R. *Biochem. J.*, 436:241–251, 2011.

[25] Telesco SE, Shih AJ, Jia F, and Radhakrishnan R. *Mol. Biosyst.*, 7:2066–2080, 2011.

[26] Telesco SE and Radhakrishnan R. *Biophys. J.*, 96(6):2321–2334, 2009.

[27] Yun C-H, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong K-K, Meyerson M, and Eck MJ. *Proc. Natl. Acad. Sci. USA*, 105(6):2070, 2008.

[28] Zhang X, Gureasko J, Shen K, Cole PA, and Kuriyan J. *Cell*, 125(6):1137, 2006.

[29] Zhang X, Pickin KA, Bose R, Jura N, Cole PA, and Kuriyan J. *Nature*, 450(7170):741, 2007.

[30] Kastner J, Loeffler HH, Roberts SK, Martin-Fernandez ML, and Winn MD. *Journal of Structural Biology*, 167(2):117, 2009.

[31] Martin-Fernandez ML. *Biochemical Society Transactions*, 40:184, 2012.

[32] Roberts SK, Tynan CJ, Winn M, and Martin-Fernandez ML. *Biochemical Society Transactions*, 40:189, 2012.

[33] Bagossi P, Horváth G, Vereb G, Szöllösi J, and Tözsér J. *Biophys J.*, 88(2):1354–1363, 2005.

[34] Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, Kuriyan J, and Shaw DE. *Cell*, 152(3):557, 2013.

[35] Marrink SJ and Tieleman DP. *Chem. Soc. Rev.*, 42(16):6801, 2013.

[36] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, and de Vries AH. *Journal of Physical Chemistry B*, 111(27):7812, 2007.

[37] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, and Marrink SJ. *Journal of Chemical Theory and Computation*, 4(5):819, 2008.

[38] Hall BA, Chetwynd AP, and Sansom MSP. *Biophysical Journal*, 100(8):1940, 2011.

[39] Khalfa A and Tarek M. *Journal of Physical Chemistry B*, 114(8):2676, 2010.

[40] Marrink SJ, de Vries AH, and Tieleman DP. *Biochimica Et Biophysica Acta-Biomembranes*, 1788(1):149, 2009.

[41] JF Franco-Gonzalez, J Ramos, VL Cruz, and J Martínez-Salazar. *J. Mol. Model.*, 19(3):1227–1236, 2013.

[42] Franco-Gonzalez JF, Ramos J, Cruz VL, and Martinez-Salazar J. *J. Mol. Model.*, 19(2):931–941, 2013.

[43] Vicente-Alique E, Nunez-Ramirez R, Francisco Vega J, Hu P, and Martinez-Salazar J. *European Biophysics Journal with Biophysics Letters*, 40(7):835, 2011.

[44] Burgess AW, Cho HS, Eigenbrot C, Ferguson KM, Garrett TPJ, Leahy DJ, Lemmon MA, Sliwkowski MX, Ward CW, and Yokoyama S. *Molecular Cell*, 12(3):541, 2003.

[45] Ghosh R, Narasanna A, Wang SE, Liu S, Chakrabarty A, Balko JM, Gonzalez-Angulo AM, Mills GB, Penuel E, Winslow J, Sperinde J, Dua R, Pidaparthi S, Mukherjee A, Leitzel K, Kostler WJ, Lipton A, Bates M, and Arteaga CL. *Cancer Research*, 71(5):1871, 2011.

[46] Periole X, Cavalli M, Marrink S-J, and Ceruso MA. *Journal of Chemical Theory and Computation*, 5(9):2531, 2009.

[47] Fuentes G, Scaltriti M, Baselga J, and Verma CS. *Breast Cancer Res.*, 13:1–9, 2011.

[48] Hess B, Kutzner C, van der Spoel D, and Lindahl E. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[49] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[50] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[51] Hess B, Bekker H, Berendsen HJC, and Fraaije JGEM. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[52] Marrink SJ, de Vries AH, and Mark AE. *Journal of Physical Chemistry B*, 108(2):750, 2004.

[53] Amadei A, Linssen ABM, and Berendsen HJC. *Proteins Struct. Funct. Bioinforma.*, 17(4):412–425, 1993.

[54] Amadei A, Linssen AB, de Groot BL, van Aalten DM, and Berendsen HJ. *J. Biomol. Struct. Dyn.*, 13(4):615–625, 1996.

[55] Siuda I and Thogersen L. *Journal of Molecular Modeling*, 19(11):4931, 2013.

[56] Webb SED, Roberts SK, Needham SR, Tynan CJ, Rolfe DJ, Winn MD, Clarke DT, Barraclough R, and Martin-Fernandez ML. *Biophysical Journal*, 94(3):803, 2008.

[57] Lu C, Mi n Z, Grey MJ, Zhu J, Graef E, Yokoyama S, and Springer TA. *Mol. Cell Biol.*, 30(22):5432–5443, 2010.

[58] Endres NF, Das R, Smith AW, Arkhipov A, Kovacs E, Huang Y, Pelton JG, Shan Y, Shaw DE, Wemmer DE, Groves JT, and Kuriyan J. *Cell*, 152(3):543, 2013.

[59] Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, Wemmer DE, Zhang X, and Kuriyan J. *Cell*, 137(7):1293, 2009.

[60] Choowongkomon K, Carlin CR, and Sonnichsen FD. *Journal of Biological Chemistry*, 280(25):24043, 2005.

Concluding Remarks

The present thesis yields a useful collection of homology models for the ECD ErbB receptors updated with the information provided by the latest crystallographic structures of suitable templates deposited in the Protein Data Bank. In particular, the model for ErbB3 and ErbB4 receptors includes the structure of subdomain IV which was absent previously. These models were subsequently refined by MD simulation. A set of common features was found for all the receptors, namely a *periscope* movement of the dimerization arm in subdomain II. What is more important is the remarkable flexibility found for subdomain IV. A *hinge* movement of this domain towards subdomains II and III was observed in all cases. A model for the interaction of ErbB2 and ErbB3 was also proposed. This complex forms one of the most biologically relevant dimers associated with aggressive carcinomas. To the best of our knowledge, the model proposed is the first atomistic scale model for ECD interaction in this heterodimer. The *hinge* movement observed in the separated receptors is also noticeable in the complex in an asymmetric way, being less mobile in ErbB2 than in ErbB3. The biological consequences of this information are not so evident, and further studies need to be carried out. However, the generation of the ECD ErbBs models presented in this thesis will serve as an starting point for a systematic study of these important receptors in order to provide clues for the development of more effective therapeutic strategies.

MD and PCA studies of the flexibility of the complex between the extracellular domain of the ErbB2 receptor and the Trastuzumab antibody has been presented. Both secondary structure for the complex and putative interactions between subdomain IV of ErbB2 and variable domains (V_H and V_L) of Trastuzumab are well conserved along the MD trajectory. Besides that, a *hinge* move approaching the subdomain II to the Trastuzumab component is revealed by combining the MD trajectory and PCA analysis of the most relevant eigenvectors. This collective motion allows the interaction between the dimerization arm of the ECD-ErbB2 subdomain II and subdomain C_H of the antibody. In addition, it has been observed some differences attributed to the crystal packing between the

MD simulation and the X-ray structure. Thus, the monomer packing in the crystalline cell hinders the *hinge* move discussed above due to the presence of two other nearby ECD-ErbB2/Trastuzumab-Fab complexes, in this manner preventing the interaction between the dimerization arm and the C_H subdomain of its own Trastuzumab protein. The observed interaction of the antibody with the dimerization arm could provide new clues to design possible modifications of the antibody that may then enhance its therapeutic power. An effective blockade of the dimerization arm would disrupt the ErbB dimerization and consequently would lead to the interruption of the signaling cascade. The simultaneous effect on both ErbB2 subdomains II and IV exerted by a modified Trastuzumab would be of great interest in the treatment of ErbB2 over-expressed tumors.

A comparative study between the Martini CG and the atomistic OPLS-AA models applied to the MD simulation of the ECD-ErbB2/Trastuzumab-Fab have been performed. In particular, the following CG settings have been evaluated: the effect of the cutoff radius to find neighbors (r_{list}), the influence of the method to calculate the electrostatic interactions (PME or Shift) and the usage the Elastic Networks either by domains or just for the whole protein (domElNeDyn or ElNeDyn). In general the CG simulations are able to satisfactorily reproduce the crystallographic determined interaction contacts between the ECD-ErbB2 and Trastuzumab-Fab proteins as well as the additional interface involving the ErbB2 subdomain II already described in the atomistic simulations. The PME treatment of electrostatic interactions seems to behave as good as or even better than the Shift methodology in view of the different results reported in this thesis. The ElNeDyn protocol for the treatment of elastic networks shows more flexibility when the networks are built by domains instead of considering the whole protein. It allows exploring a wider region of the conformational space at the expense of some mismatch with the atomistic models. It also can be concluded that the larger neighbor lists cut-off radius of 1.4 nm gives better results when compared with the atomistic system. The combination of PME, r_{list} of the 1.4 nm and an EN by domains gives an overall satisfactory correlation with the atomistic simulation structures, in particular, the protein-protein interface has been captured adequately by the CG model. This result can have important consequences regarding the consideration of using GPU technologies. At present, the GPU version of the Gromacs package used for MD simulations contemplates only the usage of PME for electrostatic interaction calculation. The GPU technology yields supercomputer performance on common server machines, so it is receiving considerable attention by software developers.

Taking into account the results mentioned above, for the first time, MD simulations of full-length models of ErbB2 monomer and dimer systems embedded in a lipid bilayer and complexed with the Trastuzumab-Fab monoclonal antibody have been reported. The system has been simulated within the CG approach using the Martini force field, based on the parameters prevously established. The

results presented in this thesis show a satisfactory comparison with experimental data and atomistic simulations. It has been carried out several microsecond CG MD simulations on systems containing either the full-length Erbb2 monomer or dimer with or without the monoclonal antibody Fab and embedded in a lipid bilayer. The model proposed in this thesis seems to fulfill a number of experimental and computational observations about the full-length ErbB2 structure and its interaction with the lipid bilayer. Some details observed in atomistic simulations of the different ErbB domains are reproduced in the CG models, as for example: *(i)* in the monomer case an interaction between the antibody constant domain and the dimerization arm can be observed in addition to the native interaction with the epitope in subdomain IV; *(ii)* the ErbB2 receptor extracellular domain tends to lay down on the lipid bilayer surface in agreement with recent FRET analysis and detailed atomistic simulations of homolog systems; *(iii)* the intracellular domain also presents a noticeable interaction with the bilayer surface; and, *(iv)* the JM fragment has been found to partially penetrate the membrane bilayer, in agreement with NMR measurements made with this fragment in a micelle environment. On the other hand, the inclusion of the Trastuzumab Fab monoclonal antibody has a pronounced effect on the structural disposition of the ectodomain, due to the hindrance exerted on those amino acids mainly responsible for the interaction at subdomain IV C-terminal. The antibody containing system shows a displacement towards the C-terminal segment of the transmembrane domain contacts in the dimer and a complete lost of the JM antiparallel arrangement characteristic of the active state that is also observed in the pure homodimer case. This action mechanism may explain the lost of signaling ability impaired by Trastuzumab to ErbB2 homodimers experimentally found.

Conclusiones Finales

La presente tesis aporta una colección muy útil de modelos de homología de los dominios ECD de los receptores ErbB, actualizada con la información proveista por las últimas estructuras cristalográficas de patrones apropiados depositados en la base de datos Protein Data Bank (PDB). En particular, el modelo para los receptores ErbB3 y ErbB4 incluyen las estructuras del subdominio IV el cual estaba ausente previamente. Estos modelos fueron subsecuentemente refinados por simulaciones de MD. Se ha encontrado un conjunto de características comunes para todos los receptores, concretamente un movimiento de *periscopio* del brazo de dimerización en el el subdominio II. Lo que es aún más importante es la notable flexibilidad encontrada para el subdominio IV. En todos los casos se ha observado un movimiento de *bisagra* de este dominio hacia subdominios II y III. También ha sido pro-puesto un modelo para la interacción de la ErbB2 y ErbB3. Este complejo forma uno de los dímeros más relevantes asociados con carcinomas agresivos. Hasta donde tenemos conocimiento, el modelo propuesto es el primer mo-delo a escala atomística para la interacción ECD en este heterodímero. El movimiento de *bisagra* observado en los receptores de manera separada es también notable en el complejo en una forma asimétrica, siendo menos móvil la ErbB2 que la ErbB3. Las consecuencias biológicas de esta información no son muy evidentes, por lo tanto es necesario llevar a cabo futuros estudios. Sin embargo, la generación de modelos para los ECD de ErbBs presentados en esta tesis servirán como un punto de partida para un estudio sistemático de estos importantes receptores con el fin de proveer pistas para el desarrollo de estrategias terapéticas más efectivas.

Por otro lado, se han presentado estudios de MD y PCA de la flexibilidad del complejo entre el dominio extracelular del receptor ErbB2 y el anticuerpo Trastuzumab. Tanto la estructura secundaria para el compejo como las presuntas interacciones entre el subdominio IV de la ErbB2 y el dominio variable (V_H y V_L) del Trastuzumab son bien conservadas a lo largo de la trayectoria de MD. Además, combinando MD y los autovectores más relevantes del análisis de PCA se revela un movimiento de *bisagra* que aproxima el subdominio II al Trastuzumab. Estos movimientos colectivos permiten la interacción entre el brazo de dimerización del ECD-ErbB2 en el subdominio II y el subdominio C_H del anticuerpo.

Adicionalmente, se han observado algunas diferencias atribuidas al empaquetamiento cristalino entre la simulación MD y la estructura de Rayos-X. Así, el empaquetamiento del monómero en la celda cristalina dificulta el movimiento *bisagra* discutido anteriormente debido a la presencia de otros dos complejos vecinos de ECD-ErbB2/Trastuzumab-Fab, de esta manera se impide la interacción entre el brazo dimerización y el subdominio C_H de su propio anticuerpo Trastuzumab. La interacción observada del anticuerpo con el brazo de dimerización podría ofrecer nuevas pistas para diseñar posibles modificaciones del anticuerpo que puedan mejorar su poder terapéutico. Un bloqueo efectivo del brazo de dimerización podría afectar la dimerazación de ErbBs y consecuentemente permitiría la interrupción de la cascada de señalización. Los efectos simultáneos en sendos subdominios de la ErbB2, II y IV ejercidos por un Trastuzumab modificado podría ser de gran interés en el tratamiento de tumores por sobre-expresión de ErbB2.

Complementariamente, se ha realizado un estudio comparativo entre los modelos de CG Martini y AA OPLS aplicados a la simulación MD del complejo ECD-ErbB2 / Trastuzumab-Fab. En particular, se han evaluado los siguientes parámetros en los algoritmos de MD para simulaciones CG: el efecto del radio *cut-off* para encontrar vecinos (r_{list}), la influencia del método para calcular las interacciones electrostáticas (PME o Shift) y el uso de las redes elásticas, ya sea por dominios o sólo para toda la proteína (domElNeDyn o ElNeDyn). En general, las simulaciones CG son capaces de reproducir satisfactoriamente los contactos de interacción determinados cristalográficamente entre las proteínas ECD-ErbB2 y el Trastuzumab-Fab, así como la interface de interacción adicional que involucra el subdominio II de ErbB2 ya descrita en las simulaciones atomísticas. El tratamiento PME para las interacciones electrostáticas parece comportarse tan bien o incluso mejor que el método Shift a la luz de los diferentes resultados presentados en esta tesis. El protocolo ElNeDyn para el tratamiento de redes elásticas muestra más flexibilidad cuando las redes se construyen por los dominios en lugar de considerar la proteína entera. Esto permite la exploración de una región más amplia del espacio conformacional a expensas de alguna falta de coincidencia con los modelos atómicos. También se puede concluir que a radios *cut-off* de las listas de vecinos más grandes de 1.4 nm dan mejores resultados que radios menores cuando es comparado con el sistema atomístico. La combinación de PME, r_{list} de 1.4 nm y una EN por dominios proporciona una correlación satisfactoria general con las estructuras de simulación atomística, en particular, la interface proteína–proteína ha sido capturada adecuadamente por el modelo CG. Este resultado puede tener consecuencias importantes en cuanto a la consideración de utilizar las tecnologías basadas en GPU. En la actualidad, la versión GPU del paquete Gromacs utilizado para simulaciones MD contempla sólo el uso de PME para el cálculo de la interacción electrostática. La tecnología GPU produce rendimientos de superordenador en servidores comunes, por lo que está recibiendo una considerable atención por parte de los desarrolladores de software.

Teniendo en cuenta los resultados mencionados anteriormente, se ha reportado por primera vez, simulaciones de MD de modelos enteros de sistemas monoméricos y diméricos de ErbB2 embebidos en bicapa lipídica y complejados con el anticuerpo monoclonal Trastuzumab-Fab. El sistema ha sido simulado dentro de la aproximación CG usando el campo de fuerzas Martini, basado en los parámetros previamente establecidos. Los resultados presentados en esta tesis muestran una comparación satisfactoria con datos experimentales y simulaciones atomísticas. Se han llevado a cabo varias simulaciones de CG MD con una extención de microsegundos sobre sistemas conteniendo tanto el monómero de ErbB2 entero o el dímero con o sin el anticuerpo monoclonal Trastuzumab-Fab y embebidos en la bicapa lipídica. Los modelos propuestos en esta tesis satisfacen un número de observaciones experimentales y computacionales acerca de la estructura de ErbB2 entera y su interacción con la membrana lipídica. Algunos detalles observados en simulaciones atomísticas de los diferentes dominios ErbB son reproducidos en los modelos CG, como por ejemplo: *(i)* en el caso del monómero se observa una interacción entre el dominio constante del anticuerpo y el brazo de dime-rización además de la interacción nativa con el epítopo en el subdominio IV; *(ii)* el dominio extracelular del receptor ErbB2 tiende a tumbarse en la superficie de la bicapa lipídica en concordancia con los últimos análisis de FRET y simulaciones atomísticas de sistemas homologos; *(iii)* el dominio intracelular también presenta una notable interacción con la superficie de la bicapa lipídica; *(iv)* se ha encontrado que el fragmento JM penetra parcialmente la bicapa, en acuerdo con medidas de NMR hechas con este fragmento en un ambiente micellar. Por otro lado, la inclusión del anticuerpo monoclonal Trastuzumab-Fab tiene un efecto pronunciado sobre la disposición estructural del ectodominio, debido al impedimento ejercido de aquellos aminoacidos principalmente responsables en la interacción en el subdominio IV en el C-terminal. El sistema que incluye el anticuerpo muestra un desplazamiento hacia el segmento C-terminal de los contactos del dominio transmembrana en el dímero y una completa pérdida de la disposición anti-paralela de la JM, la cual es característica del estado activo que también es observado en el caso de homodímero puro. Este mecanismo de acción puede explicar la pérdida de capacidad de señalización deteriorada por el Trastuzumab en homodímeros ErbB2.

Perspectives

Based on the knowledge learned in this thesis, a wide skyline has been uncovered that allows one to continue the research on EGFR receptors with molecular dynamics. The research on EGFR has some issues which can be described as follows: *(i)* the determination of the features that distinguish the population of cell surface EGFR that binds ligand with high affinity is a key point for a complete understanding of normal receptor activation; *(ii)* the role of ligand-independent ErbB dimerization/clustering needs to be clarified; *(iii)* full characterization of the effects of cancer mutations on ErbB activity is needed; and finally, *(iv)* greater appreciation of similarities and differences in the mechanisms of activation is also needed.

The following topics will try to describe some perspectives regarding the new ideas for a future work:

- CG models play an essential role in capturing events at larger time and spatial scales. Therefore, a study regarding the full-length ErbB receptors using coarse graining is higly recommended.
- Martini force field has been widely used on lipids in a succesfully way. Hence, taking into account the rich diversity of the membrane composition, a study of its effect on the structure-function of ErbB receptors would be prominent.
- The CG MD can be used like an enhanced sampling method. Then, a backmapping or reverse transformation work should be proposed to study the atomistic protein-protein interactions as a detailed chemical characterization.
- Mutations on ErbB receptors in the key regions could be performed to study through MD simulations.
- Protein–protein interactions are still the weakness of the power of Martini Force Field. Improvements on the potential parameters is a topic to be studied.

List of Publications

1. *Simulation of Homology Models for the Extracellular Domains (ECD) of ErbB3, ErbB4 and the ErbB2–ErbB3 Complex in their Active Conformations.* JF Franco-Gonzalez, J. Ramos, VL Cruz and J Martínez-Salazar. *Journal of Molecular Modeling*, 19(2):931-941, **2013**.
   http://dx.doi.org/10.1007/s00894-012-1613-y

2. *Conformational Flexibility of the ErbB2 Ectodomain and Trastuzumab Antibody Complex as Revealed by Molecular Dynamics and Principal Component Analysis.* JF Franco-Gonzalez, J. Ramos, VL Cruz and J Martínez-Salazar. *Journal of Molecular Modeling*, 19(3):1227-1236, **2013**.
   http://dx.doi.org/10.1007/s00894-012-1661-3

3. *Exploring the Dynamics and Interaction of a Full ErbB2 Receptor and Trastuzumab-Fab Antibody in a Lipid Bilayer Model Using Martini Coarse-Grained Force Field .* JF Franco-Gonzalez, and J. Ramos, VL Cruz and J Martínez-Salazar. *Journal of Computer-Aided Molecular Design*, 28(11):1093-1107, **2014**.
   http://dx.doi.org/10.1007/s10822-014-9787-2

4. *Protein-Protein and Protein-Membrane Interactions regarding the ErbB2/Trastuzumab-Fab Complexes. A Coarse-Grained Molecular Dynamics Description.* JF Franco-Gonzalez, J. Ramos, VL Cruz and J Martínez-Salazar. *Biophysical Journal*, 106(2):666a - 667a, **2014**.
   http://dx.doi.org/10.1016/j.bpj.2013.11.3690

5. *Comparing Atomistic and Martini Coarse-Grained Models on Protein-Protein Interactions applied to the ErbB2/Trastuzumab Complex.*
   JF Franco-Gonzalez, J. Ramos and VL Cruz. **Submitted**.